THE FIRST THREE SECONDS: A REVIEW OF POSSIBLE EXPANSION HISTORIES OF THE EARLY UNIVERSE

ROUZBEH ALLAHVERDI1, MUSTAFA A. AMIN2, ASHER BERLIN3, NICOLÁS BERNAL4, CHRISTIAN T. BYRNES5, M. STEN DELOS6, ADRIENNE L. ERICKcek6, MIGUEL ESCUDERO7, DANIEL G. FIGUEROA8, KATHERINE FRESEE9,10, TOMOHIRO HARADA11, DAN HOOPER12,13,14, DAVID I. KAISER15, TANVI KARWAL16, KAZUNORI KOHRI17,18, GORDAN KRNJAC12, MAREK LEWICKI19, KALOIAN D. LOZANOV20, VIVIAN POULIN21, KUVER SINHA22, TRISTAN L. SMITH23, TOMO TAKAHASHI24, TOMMI TENKANEN25,a, JAMES UNWIN26, VILLE VASKONEN27,28, and SCOTT WATSON28

1Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131, USA
2Department of Physics & Astronomy, Rice University, Houston, Texas 77005, USA
3Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA
4Centro de Investigaciones, Universidad Antonio Nariño, Carrera 3 Este # 47A-15, Bogotá, Colombia
5Department of Physics and Astronomy, Pevensey II Building, University of Sussex, BN1 9RH, UK
6Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Phillips Hall CB3255, Chapel Hill, North Carolina 27599, USA
7Physics Department, King’s College London, London WC2R 2LS, UK
8Instituto de Física Corpuscular (IFIC), University of Valencia-CSIC, E-46980, Valencia, Spain
9Physics Department, University of Texas, 2515 Speedway, STOP C1600, Austin, TX 78712, USA
10Oskar Klein Centre for Cosmoparticle Physics, Stockholm University, 10691 Stockholm, Sweden
11Department of Physics, Rikkyo University, Toshima, Tokyo 171-8501, Japan
12Fermi National Accelerator Laboratory, Theoretical Astrophysics Group, Batavia, IL 60510, USA
13University of Chicago, Kavli Institute for Cosmological Physics, Chicago, IL 60637, USA
14University of Chicago, Department of Astronomy and Astrophysics, Chicago, IL 60637, USA
15Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
16Center for Particle Cosmology, Department of Physics & Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
17Theory Center, IPNS, KEK, and Sokendai, Tsukuba, Ibaraki 305-0801, Japan
18Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
19Faculty of Physics, University of Warsaw ul. Pasteura 5, 02-093 Warsaw, Poland
20Max Planck Institute for Astrophysics, Karl-Schwarzschild Str. 1, Garching, 85748, Germany
21Laboratoire Univers & Particules de Montpellier (LUPM), CNRS & Université de Montpellier (UMR-5299), Place Eugène Bataillon, F-34095 Montpellier Cedex 05, France
22Department of Physics and Astronomy, University of Oklahoma, Norman, OK, 73019, USA
23Department of Physics and Astronomy, Swarthmore College, 500 College Ave., Swarthmore, PA 19081, USA
24Department of Physics, Saga University, Saga 840-8502, Japan
25Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA
26Department of Physics, University of Illinois at Chicago, 845 W Taylor Street, Chicago, IL 60607, USA
27NICPB, Ràvala 10, 10143 Tallinn, Estonia and
28Department of Physics, Syracuse University, Syracuse, NY 13214, USA

ABSTRACT

It is commonly assumed that the energy density of the Universe was dominated by radiation between reheating after inflation and the onset of matter domination 54,000 years later. While the abundance of light elements is commonly assumed to be too low for a primordial nucleosynthesis of the light nuclei to dominate the Universe, it is known that the Universe was radiation dominated during Big Bang Nucleosynthesis (BBN), there is scant evidence that the Universe was radiation dominated prior to BBN. It is therefore possible that the cosmological history was more complicated, with deviations from the standard radiation domination during the earliest epochs. Indeed, several interesting proposals regarding various topics such as the generation of dark matter, matter-antimatter asymmetry, gravitational waves, primordial black holes, or dark matter halos during a nonstandard expansion phase have been recently made. In this paper, we review various possible causes and consequences of deviations from radiation domination in the early Universe -- taking place either before or after BBN -- and the constraints on them, as they have been discussed in the literature during the recent years.

a Corresponding author
ttenkan1@jhu.edu
ville.vaskonen@kcl.ac.uk

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.
## Contents

1. Introduction 3

2. Causes of nonstandard expansion phases 5
   2.1. Post-inflation reheating (Authors: M. A. Amin, D. I. Kaiser & K. D. Lozanov) 5  
      2.1.1. Single-field phenomena 5  
      2.1.2. Multifield phenomena 6  
      2.1.3. Future directions 7  
   2.2. Extra stages of inflation (Authors: T. Tenkanen & V. Vaskonen) 8  
      2.2.1. Multistep inflation 8  
      2.2.2. Thermal inflation 9  
   2.3. Heavy particles and dark sectors (Authors: N. Bernal & J. Unwin) 9  
      2.3.1. Nonstandard cosmology from heavy particles 9  
      2.3.2. The transition to radiation domination 10  
      2.3.3. Nonstandard cosmology beyond heavy particles 11  
   2.4. Moduli fields (Authors: K. Sinha & S. Watson) 12  
   2.5. Post-BBN modifications (Authors: T. Karwal & T. L. Smith) 14

3. Consequences of nonstandard expansion phases 16  
   3.1. Consequences for dark matter (Authors: A. Berlin, D. Hooper & G. Krnjaic) 16  
   3.2. Phase transitions and baryon asymmetry (Authors: R. Allahverdi & M. Lewicki) 19  
      3.2.1. Phase transition dynamics 19  
      3.2.2. Electroweak baryogenesis 19  
      3.2.3. Constraints on baryogenesis from dilution by reheating 20  
   3.3. Consequences for models of inflation (Authors: K. Freese & T. Tenkanen) 21  
   3.4. Curvaton scenario (Authors: C. Byrnes & T. Takahashi) 24  
      3.4.1. Curvaton perturbations and local non-Gaussianity 24  
      3.4.2. Isocurvature perturbations and the evolution of non-Gaussianity after inflation 26  
   3.5. Formation of microhalos (Author: A. Erickcek) 27  
      3.5.1. Growth of density perturbations 27  
      3.5.2. Cutting off microhalo formation 28  
      3.5.3. The microhalo population 29  
   3.6. Formation of primordial black holes (Authors: T. Harada & K. Kohri) 30  
      3.6.1. Radiation-dominated universe 30  
      3.6.2. Matter-dominated universe 31

4. Constraints on nonstandard expansion phases 32  
   4.1. BBN and CMB constraints (Authors: M. Escudero & V. Poulin) 32  
      4.1.1. A brief review of standard BBN 32  
      4.1.2. BBN constraints on nonstandard cosmologies 33  
      4.1.3. Effects of nonstandard expansion rate on the CMB 34  
      4.1.4. Combining CMB with BAO and SN1a 36  
      4.1.5. Probing perturbation dynamics with CMB 37  
   4.2. Observational probes of microhalos (Authors: M. S. Delos & A. L. Erickcek) 38  
      4.2.1. Gravitational signatures 38  
      4.2.2. Dark matter annihilation within microhalos 38  
      4.2.3. The microhalo population within galaxies 39  
   4.3. Observational probes on primordial black holes (Authors: T. Harada & K. Kohri) 40  
      4.3.1. PBH evaporation and its consequences 40  
      4.3.2. Observational constraints on PBHs 41  
      4.3.3. Constraints on the power spectrum of curvature perturbation 43  
      4.3.4. Constraints on curvature perturbations from PBH formation during matter dominance 43  
   4.4. Gravitational wave backgrounds (Authors: D. G. Figneroa & M. Lewicki) 45  
      4.4.1. Inflation 45  
      4.4.2. Cosmic strings 47  
      4.4.3. Phase transitions 48

5. Summary 49
1. INTRODUCTION

More than forty years ago, Steven Weinberg famously summarized the state of early-universe cosmology in his book, *The First Three Minutes* [1]. Weinberg began his account by describing an early time when the Universe was filled with radiation. Since Weinberg’s book first appeared, cosmologists have made enormous progress in understanding the physics of the very early Universe. Yet there remains a gap in our understanding of cosmic history—a gap that spans the first few seconds. How, during that brief time interval, did the energy density of the universe come to be dominated by relativistic particles? In other words, what set the conditions at the start of Weinberg’s analysis?

The earliest epoch we can confidently gain information about is inflation, an early period of accelerated expansion in the very early Universe [2–7], which makes predictions consistent with high-precision measurements of the temperature anisotropies in the Cosmic Microwave Background (CMB) [10–14]. These observations limit the energy scale during inflation to be smaller than $O(10^{16})$ GeV, well below the Planck scale $M_P = 1.22 \times 10^{19}$ GeV. Meanwhile, constraints on the effects of neutrinos on the CMB, as well as Big Bang Nucleosynthesis (BBN) constraints, reveal that the Universe had to be radiation dominated (RD) by the time neutrinos decoupled from the thermal plasma, at the energy scale $O(1)$ MeV [15–19]. Between these two epochs stretches a period of cosmic history that is not well constrained by observations (see Fig. 1). Although the time interval between the end of inflation and neutrino decoupling is small ($\sim 1$ sec), the energy scale can drop during this period by almost 20 orders of magnitude and the Universe may have expanded by up to 60 e-folds [20–22].

The rate at which the Universe expands depends on the (spatially averaged) total energy density $\langle \rho \rangle$ and pressure $\langle p \rangle$ of the combination of fluids that fill the Universe. In particular, in a spatially flat Friedmann-Lemaître-Robertson-Walker (FLRW) Universe filled with components that can be modelled as perfect fluids, the Einstein field equations imply that the scale factor $a(t)$ evolves as

$$a(t) \propto t^{\frac{2}{3(1+w)}}.$$

where $t$ is cosmic time and the equation of state parameter, $w$, is given by

$$w = \frac{\langle p \rangle}{\langle \rho \rangle}.$$

During inflation the equation of state must be $w \approx -1$, whereas RD expansion corresponds to $w \approx 1/3$. Somehow the Universe must have transitioned from $w \approx -1$ to $w \approx 1/3$ between the end of inflation and neutrino decoupling.

According to standard cosmology, after an initial inflationary phase the Universe quickly entered a period of radiation domination, followed by an era of matter domination, and finally transitioned to the dark energy dominated period which we reside in today. However, this simple four stage evolution of the Universe need not necessarily be the case. Indeed, at the moment we do not have much observational information about the state of the Universe prior to BBN that occurred during radiation dominance at temperature $T \gtrsim 5$ MeV [23–29]. It is therefore possible that the cosmological history featured, for example, an additional early matter-dominated epoch (EMDE) due to e.g. slow post-inflationary reheating or massive metastable particles that dominated the total energy density. Potentially more exotic scenarios that change the expansion history of the Universe compared to the standard RD case can also be realized, and they can have important consequences for the physics of the early Universe.

BBN is a cornerstone of the standard cosmological model. At present, the primordial abundances of helium and deuterium are measured with 1% precision. This makes BBN a powerful arena to test nonstandard cosmological expansion histories. A common lore is that the Universe must be radiation dominated prior to BBN. This is due to the fact that the abundances of light elements are extremely sensitive to the expansion rate, thermalization of neutrino background, and the time-evolution of the neutron-to-proton ratio and their measured values agree very well

![Fig. 1. — The evolution of the comoving Hubble scale $1/(aH)$ as a function of the scale factor $a$. The expansion history between inflation and the subsequent, post-BBN era remains unknown. Here $a_{end}$ marks the end of inflation, $a_{BBN}$ denotes the commencement of BBN, and $a_{eq}$ is the time of matter-radiation equality at redshift $z \approx 3400$.](image-url)

1 For alternatives to cosmic inflation, see e.g. Refs. [8, 9].
with the predictions obtained assuming standard RD expansion. Likewise, measurements of the CMB can be used to achieve information about the expansion history of the early Universe. However, the standard flat ΛCDM model, while successful at explaining numerous cosmological data sets, currently faces several potential challenges. There are theoretical motivations to consider physics beyond ΛCDM, such as the unknown nature of inflation, dark matter (DM), dark energy, and mechanisms to explain neutrino masses. As observations have become increasingly precise, several anomalies have appeared, including the Hubble tension [30], the $σ_8$ tension [31], the $A_{\text{Lens}}$ anomaly [32, 33], the coincidence problem [34, 35], the small-scale structure problem [36], and the anomalous measurement by EDGES [37]. Theorists are hence motivated in multiple ways to postulate new physics beyond ΛCDM.

Notably, early periods of nonstandard cosmology can alter the cosmological abundances of particle species. In particular, they can have a significant impact on our expectations for the DM relic abundance, since nonstandard eras can change the expansion rate of the Universe [38, 39], lead to entropy injections that may dilute the DM relic abundance [40], or provide a nonthermal production mechanism for DM [41, 42]. Furthermore, a period of nonstandard cosmology may also impact small scale structure formation, potentially enhancing or impeding the formation of dense ultracompact DM microhalos [43–47], axion miniclusters [48, 49], and primordial black holes (PBH)s [50–56]. The survival of nongravitational relics (including, for example and in addition to DM, matter-antimatter asymmetry [57–61], primordial magnetic fields [62–66], or effective number of relativistic species [67, 68]) will depend on the details of the expansion history and post-inflationary thermalization of radiation. Along with large-scale gravitational effects associated with the expansion history, smaller scale gravitational effects – including horizon-scaled variations of expansion history and associated non-Gaussianities [69–78], imprints from small-scale clustering of matter [43, 79, 80], PBHs [81–93], and gravitational waves resulting from violent nonlinear dynamics [94–105] – provide hope for future observational probes of the time before BBN.
All of the above aspects are topics of this paper. In the following we review various possible causes and consequences of deviations from radiation domination in the early Universe – taking place either before or after BBN – and the constraints on them, as they have been discussed in the literature during the recent years. We emphasize that this paper is not an introduction to the topic and the underlying physics but a review of recent advances related to nonstandard cosmological epochs. Several interesting proposals, regarding for example the generation of DM or matter-antimatter asymmetry during a nonstandard expansion phase, warrant further research. We hope that this review helps in guiding this endeavour.

The paper is organized into three sections: Sec. 2 concentrates on the causes, Sec. 3 on the consequences, and Sec. 4 on the constraints on nonstandard expansion eras. Each section is further divided into subsections, which discuss the general topic of the section in more detail from both the model-building as well as observational point of view. Finally, in Sec. 5, we summarize our discussion. The paper contains various quantities and abbreviations that are used throughout the paper. They are laid out in Table 1.

2. CAUSES OF NONSTANDARD EXPANSION PHASES

In this section, we will discuss several causes of nonstandard expansion eras, taking place either before or after BBN. Such causes of nonstandard expansion can broadly be grouped as (i) modifications to the properties of the standard, CDM, components, and (ii) the introduction of new components altogether. These two categories can be further refined into modifications to the background and/or perturbative evolution. These categories are not exclusive, and some models may fall within multiple descriptions, as we will discuss.

2.1. Post-inflation reheating (Authors: M. A. Amin, D. I. Kaiser & K. D. Lozanov)

In this section, we focus on the period right after inflation. For reviews, see Refs. [106–110]. In the simplest models of inflation consistent with observations [12–14], inflation is driven by a slowly rolling scalar field with a potential $V(\phi) \propto \phi^n$. The shape of the potential affects the characteristics of the primordial perturbations which ultimately seed the CMB anisotropies; for values of the inflaton field relevant to scales probed by the CMB, observations constrain the potential to be more shallow than quadratic ($\alpha < 2$). The dynamics at the end of inflation, on the other hand, are determined by the shape of the potential near its minimum, as well as by the inflaton’s couplings to other fields, including SM fields and/or intermediaries (see Fig. 2).

2.1.1. Single-field phenomena

We first consider typical effects in single-field models that have negligible couplings to other fields. As shown in Fig. 2, at large field values the potential takes the form $V(\phi) \propto \phi^n$ with $n < 2$ (to be consistent with observations), near the minimum we assume a power-law of the form $V(\phi) \propto |\phi|^{2n}$, and we assume that the potential interpolates smoothly between these two forms at an intermediate scale $|\phi| \sim M$.

During inflation, the inflaton field $\phi(t, x) = \varphi(t) + \delta \phi(t, x)$ will be dominated by the spatially homogeneous component, with $|\delta\phi|/|\phi| \ll 1$. After slow-roll evolution has ended, the homogeneous condensate $\varphi(t)$ will oscillate around $t$

\[
\phi^{<2} \sim M
\]

\[
V(\phi) \propto |\phi|^{2n}
\]

\[
\delta\phi, \chi_i, \psi_j, A^\nu_\mu, g_{\mu\nu}
\]

\[
\text{Standard Model + Dark Matter}
\]

Fig. 2.— The energy stored in the almost homogeneous inflaton is transferred to its own perturbations, as well as other fields, and eventually to the SM fields and DM.
inflaton field minimally coupled to gravity (and neglecting coupled metric perturbations), the equation of motion for Fourier modes of the inflaton fluctuations takes the form

$$\ddot{\phi}_k + 3H\dot{\phi}_k + \left[ \frac{k^2}{a^2} + V_{,\phi}(t) \right] \phi_k = 0,$$

(4)

where overdots denote derivatives with respect to cosmic time $t$, $H \equiv \dot{a}/a$, and $k$ is comoving wavenumber. As $\varphi(t)$ oscillates quasi-periodically at the end of inflation, the modes $\phi_k(t)$ behave like oscillators with time-varying frequencies. In the limit $|V_{,\phi}| \gg H^2$, solutions to Eq. (4) generically take the form

$$\phi_k(t) = \mathcal{P}^+_k(t) e^{i\mu_k t} + \mathcal{P}^-_k(t) e^{-\mu_k t},$$

(5)

where $\mathcal{P}^\pm_k(t) = \mathcal{P}^\pm_k(t + \tau)$, with $\tau$ the period of oscillation of $V_{,\phi}(t)$. The quantity $\mu_k$ is known as the Floquet exponent. It is generally a complex function of $k$, $\tau$, and the amplitude of $V_{,\phi}(t)$. Modes $\phi_k(t)$ whose wavenumbers $k$ lie within so-called “resonance bands” $\Delta k$, for which $\text{Re}(\mu_k) \neq 0$, will become exponentially amplified due to parametric resonance. Such exponential growth corresponds to a rapid, nonadiabatic transfer of energy from the homogeneous condensate into higher-momentum inflaton particles. Moreover, the resonance process can involve the collective decay of multiple inflatons from the condensate into higher-momentum particles, making the process intrinsically nonperturbative.

The resonant amplification of inhomogeneities will fragment the homogeneous condensate, and the equation of state will approach that of radiation, $w = 1/3$, even in the absence of couplings to other fields [115]. The number of e-folds between the end of inflation and the beginning of RD epoch is given by [115, 116]

$$\Delta N_{\text{rad}} \gtrsim \frac{n + 1}{3} \ln \left[ \frac{k}{\Delta k} \frac{\sqrt{8\pi} M |4 - 2n|}{n + 1} \right] \quad n \neq 0, 2,$$

(6)

where $\Delta k/k$ is the relative width of the narrow resonance band. Eq. (6) holds for $n \neq 0, 2$ and $M \lesssim M_P$, and yields $O(1) \lesssim \Delta N_{\text{rad}} \lesssim O(10)$. For $n = 2$, Eq. (3) indicates that the homogeneous evolution already has $w = 1/3$. In that case, if $M \ll M_P$, then $\Delta N_{\text{rad}} \lesssim 1$ due to broad self-resonance. If the inflaton field couples to other light fields, the duration to RD epoch will be shortened. In this sense, Eq. (6) provides an upper bound on the duration between the end of inflation and onset of RD epoch.

For single-field models with $n = 1$, Eq. (3) indicates that the equation of state remains $w = 0$ with or without fragmentation of the homogeneous condensate. Even in this case, the post-inflationary epoch may be characterized by interesting nonlinear physics. In particular, for $M \ll M_P$, self-resonance can drive copious production of oscillons [116–122], which are long-lived field configurations that remain localized in space even as their amplitudes oscillate in time [123–128]. In regions of parameter space that do not yield such resonant phenomena ($M \gtrsim M_P$), the inflaton condensate will fragment due to gravitational self-interactions on a typical time-scale $\Delta N_{\text{frag}} \approx 5$ [80, 129, 130], although the equation of state of state remains $w \approx 0$. Either of these single-field scenarios with $n = 1$ would yield a long duration during which the Universe remained MD rather than transitioning to RD expansion, and hence could lead to conflicts with observational constraints similar to moduli fields discussed in Sec. 2.4.

2.1.2. Multifield phenomena

In models in which the inflaton field $\phi$ couples directly to other fields (which we may collectively denote as $\chi$), the $\chi$ fields typically acquire $\phi$-dependent effective masses, $m^2 = m^2_{\text{eff}}(\phi)$. After inflation, these couplings can drive even more efficient transfer of energy out of the inflaton condensate than in single-field models. Such fast processes can hasten the transition to $w = 1/3$, though in some cases an incomplete decay of the inflaton can yield a later phase of matter-dominated expansion, with $w \sim 0$, prior to neutrino decoupling.

For scalar fields $\chi$ that are minimally coupled to gravity, modes $\chi_k$ obey an equation of motion of the same form as Eq. (4), with $V_{,\phi} \rightarrow m^2_{\phi}(\varphi(t))$. Couplings between $\phi$ and $\chi$ are typically less constrained by observations than are self-couplings of $\phi$, and hence there can be broad resonance bands (with $\Delta k \gtrsim m_{\phi}$) within which the Floquet exponents for modes $\chi_k$ are large ($\text{Re}(\mu_k) \sim O(m_{\phi})$, where $m_{\phi}$ is the mass of the oscillating inflaton). Given the nonperturbative nature of such resonances, these decays can produce daughter particles much more massive than the inflaton [106–110, 114]. In some models (such as those involving spontaneous symmetry breaking), a tachyonic instability may occur, for which $(|k^2/\alpha^2| + m^2_k) < 0$, which likewise yields a rapid, nonperturbative amplification of certain modes $\chi_k$ [131–136].

Models with multiple scalar fields $\phi^f$ that have nonminimal couplings to gravity can have even richer resonance structure, since (upon performing a conformal transformation to the Einstein frame) the field-space manifold acquires nontrivial curvature [137]. In such models, couplings can arise between the fields from noncanonical kinetic terms as well as from direct couplings in the interaction potential. For large dimensionless nonminimal couplings $\xi_1 \gtrsim O(10^2)$, as one finds in popular models like Higgs inflation [138], these couplings can yield especially efficient resonances at
the end of inflation [139–150]. The rapid production of particles with wavenumber \( k \geq H \) can drive \( w \to 1/3 \) within \( \Delta N_{\text{rad}} \sim \mathcal{O}(1) \) after the end of inflation [151, 152].

The exponential amplification of modes driven by parametric resonance and/or tachyonic instability does not persist indefinitely. Rather, the phase of rapid growth ends due to backreaction on the inflaton oscillations from the produced particles, as well as from rescattering among modes \( \chi_k \) of different \( k \). Such backreaction effects typically grow quickly, within \( \Delta t_{\text{br}} \sim \mathcal{O}(1) \times \mu_k^{-1} \ll H^{-1} \) [106–110]. In many models, backreaction grows to be significant before resonance can completely transfer energy from the inflaton condensate into decay products, which can leave a significant fraction of the total system energy, \( \gtrsim \mathcal{O}(10\%) \), in the inflaton field.

The subsequent evolution of the equation of state is largely model dependent. For example, in models with minimally coupled scalar fields and potentials of the form \( V(\phi, \chi) = m^2 \phi^2/2 + \kappa \phi \chi^2 + g^2 \phi^2 \chi^2 \), the couplings between \( \phi \) and \( \chi \) can give rise to a long-lived, nontrivial period of expansion with \( 0 < w \lesssim 0.25 \), depending on the values of the coupling constants \( \kappa \) and \( g \) [135, 153]. At late times in such models, the expansion of the Universe will dilute the energy of relativistic decay products so much that the energy remaining in the inflaton condensate can dominate the total energy density again, giving rise to a matter-dominated phase with \( w \approx 0 \).

If the inflaton couples to vector bosons, the Universe can transition rapidly to \( w = 1/3 \) after inflation [154–161]. For example, if the inflaton is an axion-like field, then a coupling of the form \( \lambda \phi F \tilde{F} \) (where \( F_{\mu \nu} \) is the gauge-boson field-strength tensor) can lead to a virtually instantaneous transition to RD phase at the end of inflation, for sufficiently large coupling \( \lambda \) [102]. Moreover, nearly all of the energy of the inflaton condensate can be transferred to gauge fields in such processes, potentially avoiding a primordial period of expansion with \( w \simeq 0 \).

In other regions of parameter space, for sufficiently small amplitudes of the inflaton oscillations \( \varphi \) and for weak couplings \( \kappa \) and \( g \), resonances will remain subdominant, and the decay of the inflaton condensate will occur mostly due to perturbative decays. The decay rates may be estimated from Feynman diagrams for the tree-level couplings. Even for models in which strong resonances occur at early times after inflation, such perturbative decays can play an important role at late times, providing a mechanism to complete the transfer of energy from the inflaton condensate into light decay products. The rate of such perturbative decays can be enhanced in the case of bosonic fields \( \chi \). In those cases, Bose enhancement yields correction factors to the tree-level decay rates that one would estimate in vacuum, \( \Gamma_\chi \), of the form \( \Gamma_\chi \approx \Gamma_\chi^{\text{vac}} n_\chi \), where \( n_\chi \) is the occupation number. Then the change of the occupation number, \( \dot{n}_\chi = \Gamma_\chi \), can yield exponential growth rates for \( n_\chi \) rather than typical linear rates of growth in vacuum [114]. The perturbative decay of the \( \phi \) condensate will be complete when \( \Gamma_\chi^{\text{vac}} \sim H \), after which the energy will reside in relativistic decay products and the Universe will become RD.

If the inflaton couples to fermions \( \psi \), the effective mass \( m_\psi(\varphi(t)) \) can become time-dependent at the end of inflation, much as for bosonic fields \( \chi \). This can yield a very different distribution of fermionic decay products than would result from perturbative decays, although the Pauli exclusion principle prevents the occupation number for any fermionic mode to grow larger than \( 1 \) [57, 163–167]. In models in which the inflaton couples to both bosonic and fermionic fields, perturbative decays of the inflaton condensate into fermions can help complete the transfer of energy out of the inflaton and avoid a prolonged phase of \( w \simeq 0 \).

2.1.3. Future directions

The period immediately after the end of inflation can feature rich, nonlinear and nonperturbative dynamics. Parametric resonance and tachyonic instabilities can transfer energy efficiently from the inflaton condensate into higher-momentum decay products. In both single-field and multifield models, the efficient production of relativistic decay products can drive a transition in the equation of state from \( w \simeq -1 \) during inflation to \( w = 1/3 \) within a few e-folds after the end of inflation. If only slow, perturbative processes are involved, the transition can take considerably longer. Details of the energy transfer depend on the model: in some cases, the rapid transition to RD expansion can slide into a later phase of matter-dominated expansion (with \( w \simeq 0 \)) unless some mechanism exists (such as perturbative decays) to completely drain the energy from the inflaton condensate. Although the details of nonperturbative post-inflationary dynamics remain model dependent, recent work has made progress towards more model-independent means of characterizing this phase [102, 168–170].

For a given model, the full nonlinear dynamics of fields after inflation can now be explored numerically using a number of codes which evolve scalar and vector fields on a lattice [145, 171–178]. After such nonlinear evolution, however, the spectrum of decay products is typically far from an equilibrium distribution. The processes and time-scales over which such distributions of particles relax into thermal equilibrium remain much less understood, and an important area for further research; progress may be made by exploiting methods developed for the study of other systems, such as quark-gluon plasma in the context of heavy ion collisions [179]. Another interesting area of research concerns incorporating gravitational effects more fully into our understanding of post-inflation reheating [129, 177, 180–188].
Fig. 3.— The black curve shows the evolution of the Hubble horizon $1/H$ as a function of the scale factor $a$, whereas the gray dashed line shows, as an example, a scale that re-enters horizon after the matter-radiation equality.

2.2. Extra stages of inflation (Authors: T. Tenkanen & V. Vaskonen)

In this section we discuss scenarios in which the primary or primordial inflationary phase was followed by a secondary or even multiple extra stages of inflation. These additional inflationary stages may have been punctuated by an era of either standard RD expansion or by some type of nonstandard expansion; see Fig. 3 for illustration. Because the standard way of treating the dynamics of the early Universe usually contain only one period of inflation – the primordial one –, any additional stages of inflation can be seen as a nonstandard addition regardless of how the Universe expanded between the inflationary stages. The energy scales at which the additional inflationary stages may have occurred are not particularly limited, except for the usual requirements that the upper limit for the energy scale of inflation is $\rho_{\text{tot}} \simeq (10^{16}\text{ GeV})^4$ (see Sec. 3.3) and any nonstandard phase must have ended prior to BBN at $T \sim \mathcal{O}(1)\text{ MeV}$ (see Sec. 4.1).

2.2.1. Multistep inflation

We begin by discussing scenarios of multistep inflation by which we refer to models where (nonthermal) inflation occurs in at least two phenomenologically different stages with distinct characteristics, as originally suggested in Refs. [189, 190]. This type of scenarios can be generically divided into three categories: those in which the same scalar field(s) that were responsible for the primary inflation caused also the secondary stage of inflation (for instance because the inflation potential has two separate flat regions), so-called hybrid inflation models in which inflation at its last stages was driven not by the energy density of the inflaton field but by the vacuum energy density induced by an interaction to another field, and those in which another scalar field that was an energetically subdominant “spectator” field during the primary phase of inflation either acquired a large vacuum expectation value (vev) during the first inflationary stage or was kept frozen in its initial value with a finite energy density and later on became the dominant component, thus starting a new stage of inflation. Examples of the first category above include scenarios studied recently in e.g. Refs. [191–193] (see also Refs. [194, 195] for recent studies on double inflation in multifield models with a rapid turn in field-space), while scenarios belonging to the second category were originally introduced in Refs. [196, 197] and which have been studied in a large number of works since. However, generically hybrid models predict a blue-tilted spectrum for primordial perturbations\(^2\) and are therefore now ruled out by observations [13].

A secondary stage of inflation started by a spectator field was originally studied in Refs. [189, 190] and has been studied more recently in e.g. Refs. [198–212]. Typically in scenarios where the secondary stage of inflation starts due to a spectator field $\sigma$ acquiring a vev during the first stage of inflation, the vev has to be very large for the field to become energetically dominant while still in slow-roll to trigger a new period of inflation: for example, for a quadratic spectator potential $V_\sigma \propto \sigma^2$ the criterion is $|\sigma| \gtrsim M_P/\sqrt{4\pi}$. A natural mechanism for this, based on the stochastic evolution of light scalar fields in (quasi-) de Sitter space was discussed in Ref. [207]. Another possibility is based on the type of scenarios studied in e.g. Ref. [190], where the authors considered the potential

$$V = V_0 + \lambda_\phi \phi^4 \left[ \ln \left( \frac{\phi^2}{f^2} \right) - \frac{1}{2} \right] + \lambda_\sigma \sigma^4,$$  \hspace{1cm} (7)

where $f$ is a mass scale related to the ultraviolet completion of the theory. Here both fields are assumed to have a large vev initially, $\phi_i, \sigma_i > M_p$ and the field $\phi$ is assumed to drive the first stage of inflation. If $\lambda_\sigma$ is not very small, the stochastic quantum fluctuations of $\sigma$ acquired during the first stage of inflation were smaller than its initial value, and the field remained essentially frozen at $\sigma_i$. The first stage of inflation then ended due to $\phi$ rolling down to $\phi \lesssim M_P$, and the second stage began shortly after when the field $\sigma$ became the energetically dominant species. In practice, however, also a much simpler chaotic potential with a suitable choice of parameters can permit similar dynamics.\(^2\) See Eq. (38) for a parameterization of the curvature power spectrum.
2.2.2. Thermal inflation

Next, we discuss scenarios where the secondary period of inflation is caused by thermal effects. In this case the inflationary stage is known as thermal inflation [213, 214], and it is driven by a scalar field that at a high temperature is far from its zero temperature vev. Part of the total energy density of the Universe is then in the form of vacuum energy, and as temperature drops, the vacuum energy can become the dominant energy density component, therefore causing a period of inflation. Eventually, the thermal inflation ends when the scalar field either tunnels or rolls down to a lower energy minimum of its potential. Of particular interest is the former case, where the vacuum energy dominance ends with a first-order phase transition [215–218] (see also Sec. 3.2). This happens by formation of bubbles of the true vacuum which expand, finally turning the whole Universe into the new phase. These bubbles nucleate sparsely and are initially small. They then expand by many orders of magnitude and as they collide, most of the energy is in very thin regions around the bubble walls that move almost at the speed of light [219–222]. After the bubbles have collided, the scalar field oscillates around the true vacuum, potentially causing a short period of matter-dominance before decaying and reheating the radiation bath. Similarly, a short MD period can be caused in the case where the thermal inflation ends with the scalar field rolling down to the new minimum.

A period of thermal inflation can be realized, for example, in classically scale-invariant models or strongly-interacting theories [222–229]. In this case, the scalar potential is of the form

\[ V \simeq V_0 + \frac{3g^4 \phi^4}{4\pi^2} \left[ \ln \left( \frac{\phi}{v_\phi} \right) - \frac{1}{2} \right] + g^2 T^2 \phi^2, \tag{8} \]

where \( g \) is a coupling constant and \( v_\phi \) the vev of \( \phi \). The scalar field is stuck in a metastable phase that is separated from the true vacuum by a potential energy barrier generated by purely thermal effects (the last term in Eq. (8)). Therefore, a very long period of supercooling is possible, because the potential energy barrier shrinks as the temperature decreases, unlike in models where the potential energy barrier is dominated by nonthermal terms. The potential energy barrier can also disappear at a nonzero temperature in models that are not classically scale-invariant. In this case the supercooling period is typically not very long.

An interesting consequence of thermal inflation ending to a first-order phase transition is the formation of a stochastic gravitational wave background in bubble collisions (see Sec. 4.4). In addition, the collisions of very energetic bubble walls may lead to copious formation of PBHs [230–232] (see also Sec. 3.6). Alternatively, in the case that the thermal inflation ends by the field rolling down its potential PBHs can form from large density perturbations generated when the thermal inflation potential turns from convex to concave at the high-temperature minimum [233].

Finally, a secondary nonthermal stage of inflation can be caused by symmetry restoration by nonthermal effects as the particles produced in preheating after inflation can for a long time remain out of thermal equilibrium [234–236]. Also this inflationary stage ends with a symmetry breaking phase transition that can be of first order [237, 238].

A common feature of all these scenarios is that if the secondary stage of inflation lasted for a sufficiently long time, this becomes the “primordial” inflation whose imprints on the structure of the Universe at different scales are the only ones we can observe. On the other hand, if the secondary phase only lasted for a short time, the initial conditions for large and small structure formation decouple so that the first stage determines the spectrum on large scales and the second the spectrum on small scales [190]. In addition to providing initial conditions for large scale structure formation, the primary stage of inflation provides initial conditions also for all the stages that follow it. This has been argued to alleviate the problems some inflationary models face with initial conditions [191].

2.3. Heavy particles and dark sectors (Authors: N. Bernal & J. Unwin)

In this section we discuss how nonstandard cosmological histories may arise from different particle physics scenarios including heavy particles or dark sectors.

2.3.1. Nonstandard cosmology from heavy particles

We start with the intuitive idea that a long lived heavy particle \( \phi \) could source an EMD period. The state \( \phi \) could initially be in thermal equilibrium with the SM states, or could be out-of-equilibrium at all times. Once \( \phi \) becomes nonrelativistic its contribution \( \rho_\phi \) to the energy density of the Universe will grow relative to the contribution due to SM radiation \( \rho_R \). Their energy densities evolve with the scale factor according to \( \rho_\phi \propto a^{-3} \) while \( \rho_R \propto a^{-4} \). Thus, even if \( \phi \) initially contributes a small component to the total energy density, then provided it is sufficiently long-lived it will grow to dominate the energy density of the Universe and lead to an EMDE. These ideas first arose in the context supersymmetry and string theory, see e.g. Refs. [42, 239–242], and the particular case of string moduli fields will be discussed in Sec. 2.4.

More generally, any massive meta-stable state could potentially come to dominate the energy density of the early Universe. Specifically, in the early Universe the evolution of \( \rho_\phi \) and the SM entropy density \( s \) are governed by the
following coupled Boltzmann equations [243]

\[ \dot{\rho}_\phi + 3H\rho_\phi = -\Gamma_\phi \rho_\phi, \]  
(9)

\[ \dot{s} + 3Hs = -\frac{\Gamma_\phi \rho_\phi}{T}, \]  
(10)

where \( \Gamma_\phi \) is the \( \phi \) decay width, \( H \) is the Hubble parameter, and \( T \) is the temperature of the SM photons. Notably, the evolution of \( T \) can be tracked via Eq. (10), since the entropy density \( s(T) \equiv (\rho_R + \rho_\phi)/T = 2\pi^2 g_* s(T) T^3/45 \). Additionally, under the assumption that the SM bath maintains internal equilibrium at all times in the early Universe, its temperature dependence follows from its energy density \( \rho_R(T) = \pi^2 g_*(T) T^4/30 \). Here \( g_*(T) \) and \( g_{*s}(T) \) correspond to the effective number of relativistic degrees of freedom for SM energy and entropy densities, respectively. The Hubble expansion rate \( H \) is defined by \( H^2 = 8\pi G(\rho_R + \rho_\phi)/(3M_p^2) \). While \( \Gamma_\phi \ll H \) (such that the radiation bath is mainly composed of particles that are not produced by decays) the entropy is separately conserved in each sector and the scalar factor and temperature scale as \( T \propto a^{-1} \) (up to variations of \( g_{*s} \)). The expansion rate will be different depending on the fractional distribution of the total energy density. Specifically, during the era of MD one has \( H \propto T^{3/2} \), whereas for the RD case \( H \propto T^2 \). Also note that once \( \phi \) decays, the entropy conservation in the bath no longer holds (which corresponds to when the energy density in the relativistic decay products of \( \phi \) becomes significant), and the evolution switches to \( T \propto a^{-3/8} \) and thus \( H \propto T^{4/2} \) [38, 244] until the Universe transitions to the RD era.

Going beyond an EMD period, one can envisage instead that the Universe may have been dominated at one stage by a particle species with a general equation of state \( w \), such that \( \rho_\phi \propto a^{-3(1+w)} \) [111]. In this case the Boltzmann equation (9) depends on \( w \) and (assuming the scalar field is decaying at zero-momentum) is instead given by

\[ \dot{\rho}_\phi + 3(1+w)H\rho_\phi = -\Gamma_\phi \rho_\phi. \]  
(11)

Thus, while the \( \phi \) field dominates the energy density of the Universe the Hubble rate scales as \( H \propto a^{-2(1+w)} \propto T^{2(1+w)} \).

As in the MD case, once the energy density in the relativistic decay products of \( \phi \) becomes significant this evolution switches to \( H \propto T^4 \) (regardless of the value of \( w \) [245]). Simple cases corresponding to MD and RD expansion are found for \( w = 0 \) and \( w = 1/3 \), respectively. Examples with more general \( w \) can occur for oscillating scalar fields with certain potentials (e.g. periodic [246, 247] or special constructions [248]), in brane world cosmologies [249, 250], and for scalar-tensor theories [251, 252]. An especially well motivated case is that of kination [253–255] in which \( \phi \) is a “fast-rolling” field whose kinetic term \( \phi \) dictates the expansion rate of the post-inflation Universe, implying an equation of state \( w \approx 1 \). In most models of interest \( |w| < 1 \), with \( w = -1 \) corresponding to dark energy or quintessence domination [256, 257]. For models with \( w > 1 \) there are concerns about superluminal propagation, although it has been claimed this may be resolved in certain cases [248]. Importantly, for \( w > 1/3 \) the energy density \( \rho_\phi \) gets diluted faster than radiation, and if \( \rho_\phi \ll \rho_R \) at \( T \sim 1 \) MeV, the late decay of \( \phi \) typically has a negligible impact.

### 2.3.2. The transition to radiation domination

There are two main possibilities through which a period of nonstandard cosmology can come to an end in the case that it is sourced by a population of heavy state \( \phi \) or similar physics. Firstly, if the energy density \( \rho_\phi \) redshifts faster than radiation \( (w > 1/3) \) then after some duration the SM radiation will dominate the energy density, leading to a smooth transition to a RD Universe. Kination \( (w = 1) \) is the prime example of this scenario. Alternatively, the period of nonstandard cosmology could end due to the decays of \( \phi \) to lighter degrees of freedom which are produced relativistically. This leads to a rather more abrupt exit from the era of nonstandard expansion which coincides with \( 3H \approx \Gamma_\phi \). For scenarios with \( w < 1/3 \) (such as an EMD period) decays of \( \phi \) provide the principle route to a RD Universe prior to BBN.

A common model assumption is that the \( \phi \) states can decay to SM particles (and potentially also states beyond the SM). Note that if the decay products of \( \phi \) are produced with relativistic momenta, then when the population of \( \phi \) decay at \( H \approx \Gamma_\phi \) this causes the Universe to transition from a \( \phi \)-sourced era of nonstandard cosmology to a RD Universe. Importantly, for successful BBN the temperature at the end of the \( \rho_\phi \) dominated phase must satisfy \( T_{\text{reh}} \gtrsim 5 \text{ MeV} \) [24, 25, 28, 258, 259], where \( T_{\text{reh}} \) is determined by the decay rate, \( \Gamma_\phi = 3H(T = T_{\text{reh}}) \), giving

\[ T_{\text{reh}}^4 = \frac{5}{4\pi^3 g_\phi(T_{\text{reh}}) M_p^2 \Gamma_\phi^2}. \]  
(12)

Note that if \( T_{\text{reh}} \) is below the electroweak phase transition temperature, one must address questions regarding the origin of the baryon asymmetry; for discussion and potential resolutions see e.g. Refs. [260–263] and Sec. 3.2.

This transition to a RD Universe can have a number of important consequences, most prominently the dilution of frozen-out cosmological abundances and the nonthermal production of states, which we discuss next. For a particle species \( X \) which is decoupled from the thermal bath its comoving number density \( Y = n_X/s \) is typically expected to be constant. However, the decays of \( \phi \) can lead to changes in the comoving number density of decoupled species. The first possibility is that \( n_X \) is increased due to (nonthermal) production of \( X \) via direct \( \phi \) decays to these \( X \) states.
In particular, it has been observed that by adjusting the $\phi$ branching fraction to DM one can reproduce the observed relic density in a large range of scenarios [42, 264–276].

Alternatively, if $\phi$ decays dominantly to SM states then this leads to heating of the thermal bath relative to the decoupled species, breaking the common scaling of $n$ and $s$ [38, 243, 244, 277]. This scenario is often referred to as an “entropy injection” (or “entropy dump”) since in the instantaneous $\phi$ decay approximation it appears to imply a jump in the entropy density, although in actuality the temperature of the bath simply cools more slowly due to $\phi$ decays and at no point does the temperature increase [38]. Notably, similar to nonthermal production one can adjust the magnitude of dilution due to entropy injections to obtain the observed DM relic abundance in many scenarios. Indeed, this is the premise behind Flooded DM [262] and is also used in related scenarios which use entropy production to adjust the DM abundance after thermal freeze-out [38, 278–286]. A dedicated discussion on the consequences of nonstandard expansion phases on DM models is presented in Sec. 3.1

The magnitude of the entropy injection during the transition to RD epoch due to $\phi$ decays depends on the contribution to the total energy density from the $\phi$ states relative to the SM radiation. Assuming the instantaneous decay approximation, in which all $\phi$ states decay simultaneously at $T = T_{\text{reh}}$, the entropy prior and after decays are related via

$$s_{\text{before}} = \frac{\rho_R(T_{\text{reh}})}{\rho_\phi(T_{\text{reh}})} \frac{\rho_R(T_{\text{reh}})}{\rho_\phi(T_{\text{reh}})}^{3/4},$$

up to variations of the relativistic degrees of freedom. Instantaneous $\phi$ decays is a reasonable approximation provided no physical processes of interest (e.g. freeze-out) occur near the point of $\phi$ decays [38]. Since it is assumed that below some critical temperature $T_* \phi$ the $\phi$ energy density starts evolving as $a^{-3(1+w)}$, whereas the radiation continues to evolve as $a^{-4}$, over time the $\rho_\phi$ grows relative to $\rho_R$. Thus to calculate the entropy injection one needs to track both the initial contributions to the energy density at $T = T_*$ and the duration from this point to $T = T_{\text{reh}}$. Specifically, these energy densities can be followed from inspection of the Friedmann equation (for $H(T_*) > H(T) > a_0$)

$$H(a)^2 = H(a_*)^2 \left[ R_\phi \left( \frac{1}{\Delta a} \right)^{3(1+w)} + R_R \left( \frac{1}{\Delta a} \right)^4 \right],$$

where $\Delta a(T) \equiv a(T)/a(T_*)$ and $R_i = \rho_i/\rho_{\phi + \rho_R}|_{T = T_*}$.

Provided that $\phi$ is sufficiently long lived it that becomes the dominant energy density component, then between $T = T_*$ and $T = T_{\text{reh}}$ the scale factor grows by the following factor

$$\Delta a_\Gamma \equiv \Delta a(T_{\text{reh}}) \simeq \left[ \frac{H(a_*)^2 R_\phi}{R_\phi^2} \right]^{1/3(1+w)}.$$

If one then assumes that prior to $T = T_*$ the $\phi$ are relativistic then $\rho_\phi(T_*) = \rho_R(T_*) R_\phi/R_R$. Further supposing all the energy in $\phi$ is transferred to radiation when it decays, then it follows that the magnitude of the entropy dump is

$$\Delta s = \left[ \frac{\rho_R(T_{\text{reh}})}{\rho_\phi(T_{\text{reh}})} \right]^{3/4} \left[ \frac{\rho_R(T_*)}{\rho_\phi(T_*)} \right]^{\Delta a^{-4}} \frac{\Delta a^{-3(1+w)}}{\Delta a^{-3(1+w)}} = \left[ \frac{R_R}{R_\phi} \frac{\rho_R}{\rho_\phi} \right]^{3/4} \simeq \frac{R_R^{3/4}}{R_\phi^{3/4}} \left[ T_{\text{reh}} \right]^{1-3(1+w)} T_*. $$

This ratio is important since it implies that for decoupled particle species which are not significantly produced due to $\phi$ decays, then the associated comoving number density will be reduced by a (potentially large) factor $\Delta s$ as the states are diluted relative to the SM radiation. Note that this dilution also applies to particle asymmetries [287]. In particular, it dilutes any pre-existing baryon asymmetry in the Universe. This is discussed in detail in Sec. 3.2.

We highlight that in determining the magnitude of the entropy injection the temperature of the visible sector bath at which $\rho_\phi$ starts evolving as $a^{-3(1+w)}$, denoted $T_*$, is a critical parameter. In the simple case where $\phi$ is initially in thermal contact with the SM prior to decoupling and then becomes nonrelativistic, it is reasonable to identify $T_* \sim m_\phi$. However, if the $\phi$ decouple very early, or are never in thermal contact, the $\phi$ could well be thermally distributed with a different temperature and $T_*$ is a free parameter. We shall discuss establishing temperature difference between sectors shortly. Another important case is the case in which $\phi$ is a bosonic field $a$ oscillating in its potential, such that the average equation of state is $w \simeq 0$ (matter-like), i.e. an axion-like particle (ALP). The ALP case is distinct from the heavy particle scenario since $a$ becomes matter-like once it begins to oscillate around its minimum which occurs for ALP mass $m_a \sim H$ [288, 289], corresponding to $T_* \sim \sqrt{3} m_a M_\text{Pl}$, for a simple quadratic potential $V(a) \supset m_a^2 a^2$.}

2.3.3. Nonstandard cosmology beyond heavy particles

While our discussion thus far has been in the context of heavy elementary particles, the state responsible for determining the equation of state of the Universe need not be elementary and could be composite states, nonperturbative states [54, 290], or other forms of matter. For instance, a natural proposal is that a period of MD could arise if
the energy density of the Universe is mainly composed of PBHs which subsequently evaporate. Requiring that the Universe transitions to RD epoch prior to BBN implies an upper mass bound on the PBHs, around $6 \times 10^5$ kg [291]. Additionally, it was recently pointed out that PBHs could be very efficiently produced from the preheating instability so that they dominate the energy density of the Universe, and therefore the reheating no longer occurs because of inflaton decays, but rather due to evaporation of the PBHs [188, 292]. PBHs are discussed further in Sec. 3.6.

Dark sectors\(^3\) offer another interesting alternative through which to introduce nonstandard elements to cosmology. Although the dark and visible sectors could be in equilibrium in the early Universe, this need not necessarily be the case and in the converse scenario there can potentially be significant temperature differences between the two sectors. Such temperature differences can arise due to asymmetric inflationary reheating of sectors [293–295]. Moreover, even in the case that the visible and dark sector are initially decoupled with similar temperatures, subsequently evolution can lead to temperature differences. Examples in which decoupled dark sectors dynamically develop different temperatures include self-heating via semi-annihilation processes [296] or self-interactions [297–299], also via number-depleting annihilations via 3-to-2 [297, 300, 301] or 4-to-2 [302–305] processes, and in co-decaying DM models [306].

To appreciate (one reason) why temperature differences between sectors can be important, consider the case in which the temperature of \(\phi\) states \(T_\phi\) is significantly lower than that of the SM photon temperature \(T\). Then the \(\phi\) become nonrelativistic at \(T_\phi \sim m_\phi\) which could be much earlier than the point at which \(T \sim m_\phi\). Thus the states \(\phi\) need not necessarily be extremely heavy, relative to SM states, in order to give rise to an EMD era. Furthermore, the temperature of the visible sector at which \(\phi\) becomes nonrelativistic corresponds to \(T_v\) in Eq. (16) and, notably, if \(T_v\) can be treated as a (largely) free parameter this permits a large range in the magnitude of the entropy injection due to \(\phi\) decays, and increases the potential impact of dilution on decoupled particle species.

2.4. Moduli fields (Authors: K. Sinha & S. Watson)

String theories and M-theory compactifications yield a large number of scalar fields, or moduli. A well-studied example is Calabi-Yau compactification where these moduli correspond to the complex structure and Kähler, also below deformations of the internal manifold [307, 308]. Generating masses for moduli has been a central program in string phenomenology and cosmology [309–313]. This is necessitated by several considerations: massless scalars lead to long range interactions which are severely constrained by “fifth force” experiments. Moreover, if the couplings of moduli to different sectors of matter fields are non-universal, the corresponding long-range forces would violate the equivalence principle. An additional question is that of predictability: the vacuum expectation values of moduli determine the parameters of the low-energy theory and thus should be fixed. Moduli stabilization has been a central program in string theory over several decades and the construction of the effective potential draws upon many ingredients that are central to string theory: branes, fluxes, and strong dynamics in the hidden sector. Explicit constructions are still being actively discussed by the community [314–317].

From the perspective of low-energy phenomenology, an important question about moduli stabilization is the resulting mass \(m_\sigma\) of the modulus \(\sigma\). The physics of moduli should decouple at scales \(E \ll m_\sigma\). Nevertheless, decoupled moduli can still impact the pattern of the low-energy particle spectrum in many situations, for example when they are responsible for supersymmetry breaking [318–321]. On the other hand, the impact of light moduli on the evolution of the early Universe can also lead to many different observational predictions [267, 322–324]. In the early Universe, moduli are typically displaced from the minimum of their effective potential and undergo coherent oscillations that can mimic a MD era [268, 324]. Since moduli couple to other fields gravitationally, their decay width is given parametrically as \(\Gamma_\sigma \sim m_\sigma^3/M_5^2\). Depending on the mass \(m_\sigma\), moduli can decay late enough to affect the successes of BBN. This is the so-called cosmological moduli problem, which first appeared as the Polonyi problem in early versions of spontaneously broken supergravity. Stated most generally, the problem is that supersymmetry breaking in the hidden sector with gravity mediation often contains scalar fields that have weak scale masses and couple gravitationally. After inflation, these fields behave like nonrelativistic matter and dominate the energy density of the Universe until after BBN, spoiling its successful predictions [240, 325–327].

The cosmological moduli problem may actually be more of an opportunity if the masses of the moduli result in decays prior to BBN. Indeed, this situation leads to new cosmological predictions and an alternative to a standard thermal history. That is, if a modulus is massive enough to decay before BBN, but not so massive as to decouple entirely from low energy physics (a situation that commonly occurs in UV complete examples), its effects on DM, baryogenesis, and structure formation can lead to novel and interesting predictions [324]. The late decay of moduli can be viewed as a second reheating of the Universe (after the first, inflationary one), with the production of relativistic SM and other light (beyond the SM) particles. Depending on the masses and couplings of the modulus to other fields, such as DM and baryons, this framework can lead to new expectations for post-inflationary cosmology prior to BBN.

The critical question then is: what is the mass \(m_\sigma\) of a modulus in string theory, or, framed another way, what is the behaviour of the curvature of the effective potential near its minimum? Given that the scales relevant in string theory are the Planck scale, the string tension, and the typical size of the extra dimensions (the Kaluza-Klein scale),

---

\(^3\) It is common to refer the set of states which are charged under the SM gauge groups, or otherwise with significant couplings to SM states, as the visible sector. In contrast to this a dark sector (or hidden sector) typically refers to a set of states which are neutral singlets under the SM gauge groups and with weak direct coupling to SM states.
one would expect a priori that \( m_\sigma \) is determined by a combination of these scales and that there is no obvious role of a low-energy scale in the stabilization mechanism. If that were the case, moduli would be too massive and decay too early to leave discernible imprints on pre-BBN cosmology.

In explicitly constructed examples of moduli stabilization, a common result is that some moduli remain parametrically lighter than the string scale. This can be understood at the level of effective supergravity, since the key is the connection between moduli stabilization and supersymmetry breaking. The effective potential determines the moduli \( F \)-terms dynamically. This, combined with the Kähler metric of the matter sector, fixes the strength of gravity mediation. More concretely, the scalar potential in the supergravity limit is given by

\[
V = e^K \left( \sum_\sigma |D_\sigma W|^2 - 3|W|^2 \right),
\]

where \( W(\sigma) \) and \( K(\sigma) \) are the superpotential and Kähler potential, respectively. The gravitino mass is defined as \( m^2_{3/2} = \langle e^K |W|^2 \rangle \) and here we have set \( M_p^2/(8\pi) = 1 \). At the extremum, \( \partial_\sigma V = 0 \) and \( \partial_{\sigma\sigma} V > 0 \). Moreover, an almost vanishing cosmological constant requires \( V \simeq 0 \). Then, both terms in Eq. (17) are forced to be \( \sim m^2_{3/2} \). Since the modulus mass matrix is positive definite, the smallest eigenvalue should then be \( O(m_{3/2}) \). For a model with a single modulus with nonvanishing \( F \)-term, the trivial conclusion is \( \partial_{\sigma\sigma} V \sim m^2_{3/2} \).

This set of conditions is restrictive enough to yield \( m_\sigma \sim m_{3/2} \). Indeed, taking into account a redefinition to obtain canonical kinetic terms, one obtains \[ m_\sigma \sim \langle \sigma \rangle m_{3/2} \sim O(1 - 100) \times m_{3/2}. \] (18)

The corresponding reheat temperature is given by \( T_{\text{reh}} \sim g_*^{-1/4} \sqrt{\Gamma_{\sigma} M_p} \) or

\[
T_{\text{reh}} = e^{1/2} \left( \frac{10.75}{g_*} \right)^{1/4} \left( \frac{m_\sigma}{50 \text{ TeV}} \right)^{3/2} \text{ MeV},
\]

where \( g_* \) is the number of relativistic degrees of freedom at \( T_{\text{reh}} \), and \( e \) is a model-dependent \( O(0.1 - 1) \) number.

It is intriguing that the lower bound on the modulus mass required to avoid the cosmological moduli problem corresponds to the gravitino mass scale required to address the electroweak hierarchy problem. Another interesting result from these models is that the reheat temperature is deeply connected to fundamental physics and is not a free parameter. Since moduli are ubiquitous in string theory, a nonstandard cosmological history thus appears to be a robust possibility and perhaps an inevitable one. Of course, connecting the physics of moduli to a consistent nonstandard cosmological history presents many challenges. We now turn to a discussion of some of these issues.

One class of challenges lies in consistent embeddings of the matter sector into a given string compactification and the interactions of various particles with moduli. For example, the moduli-induced gravitino problem \[ 267, 278, 279 \] arises because gravitinos produced from moduli themselves subsequently decay and produce DM in supersymmetric models. This tends to overclose the Universe unless the decay of the modulus to the gravitino is suppressed due to branching or kinematics. Whether or not such blocking is possible is a model-dependent question. Similarly, the moduli-induced dark radiation problem arises from the fact that moduli are either accompanied by light pseudoscalars in typical string compactifications, or themselves decay to such axion-like-particles \[ 330-332 \]. The existence of such relativistic species in the dark sector is subject to \( \Delta N_{\text{eff}} \) constraints from cosmological observations, in turn imposing constraints on the moduli physics. Another class of challenges lies in various features of the low-energy effective potential of the modulus. An example of this is the “overshoot problem”: since moduli are displaced from their minima during inflation, this displacement should not be such that a modulus overcomes the barrier protecting the local low-energy minimum. This problem is again model-dependent and depends on the stabilization mechanism, and may suggest a correlation of the scales of supersymmetry breaking (the barrier height) and inflation \[ 333, 334 \]. Many of these issues require further theoretical investigation. Regardless, we have seen that a nonstandard cosmological history is a robust prediction of string/M-theoretic approaches to beyond-SM physics.

There can be a number of effects of a decaying modulus on early Universe cosmology: nonthermal production of DM \[ 329, 335 \], late-time baryogenesis \[ 261, 336, 337 \], effects on the CMB \[ 323 \] and LSS \[ 338 \], and implications for the matter power spectrum \[ 45 \], effective number of neutrino species \[ 330 \], and isocurvature constraints. We briefly discuss some of these topics, leaving a more detailed discussion for other authors in this review.

**Gravitational effects: matter power spectrum and CMB:** We first discuss the effect of moduli on the DM halo. In a RD Universe, DM perturbations \( \delta \rho / \rho \) grow logarithmically with the scale factor. In a matter-dominated Universe, on the other hand, they grow linearly: \( \delta \rho / \rho \sim a(t) \sim t^{3/2} \). This implies that while in standard cosmology the growth of structure only becomes significant after matter-radiation equality, an early MD era with cosmological moduli results in an early period of growth and a new scale at smaller wavelengths in the matter power spectrum. For \( T_{\text{reh}} \) near the BBN limit, the result may be the formation of Earth-sized, ultracompact microhalos \[ 43, 45, 339 \]. If the DM decouples both thermally and kinetically prior to the subsequent onset of RD epoch, this enhancement of
the small-scale matter power spectrum may survive, with profound consequences for WIMP detection [339–341] and axion cosmology [48]. For a more detailed discussion on effects on the matter power spectrum, see Sec. 3.5 and Sec. 4.2.

We next turn to possible effects of moduli on the CMB [232]. The main point here is that the equation of state (e.g. thermal \( w = 1/3 \) or nonthermal \( w = 0 \)) in the aftermath of inflation determines the expansion rate of the Universe. Thus, the rate at which primordial perturbations re-enter the Hubble horizon is impacted by an era when moduli dominate the energy density. The total number of e-folds is given by \( N_L = \ln(a_{end}/a_0) \), where \( a_{end} \) and \( a_0 \) are the scale factors at the end of inflation and when the pivot scale exited the horizon, respectively. It is dependent on, apart from the inflationary potential, the value of the energy density at the end of inflation and the density at which the Universe is assumed to become thermalized. The energy densities in turn depend on the mass and reheating temperature of the modulus. Thus, the properties of the modulus sector impact inflationary observables through \( N_L \) and conversely, CMB observations can be used to constrain the modulus sector. For a more detailed discussion on inflationary observables, see Sec. 3.3.

Nonthermal DM: Moduli-dominated cosmological histories have a major impact on the standard thermal freeze-out paradigm of DM. The reason is that the decay of the modulus dilutes any previously existing population of DM by a factor \( \Omega_{CDM} \rightarrow \Omega_{CDM} (T_{reh}/T_f)^3 \), where \( T_f \) is the DM freeze-out temperature. DM is produced during the modulus reheating process at low temperatures \( \sim T_{reh} \). If the number density of DM particles produced exceeds the critical value \( n^* = (H/(\sigma_{xv}))_{T= T_{reh}} \), then the particles quickly annihilate down to this value, which is an attractor. The result is a parametric enhancement of the relic density \( \Omega_{DM}^{NT} = \Omega_{DM}^{T} (T_f/T_{reh}) \). If, on the other hand, the number density of DM particles produced from modulus decay is lower than the attractor value, no further annihilation takes place and the final relic density is simply given by \( n_{DM} = n_0 \times \text{Br}_{\tau \rightarrow DM} \), where \( \text{Br}_{\tau \rightarrow DM} \) is the branching ratio for modulus decay to DM and \( n_0 \) is the number density of the scalar condensate. Due to the extra freedom in choosing \( T_{reh} \) in the first case and \( \text{Br}_{\tau \rightarrow DM} \) and \( n_0 \) in the second, a much wider range of DM masses and annihilation cross-sections can satisfy the observed relic density constraint compared to the thermal freeze-out scenario. For a more detailed discussion on DM, see Sec. 2.3 and Sec. 3.1.

Baryogenesis: Just as in the case of DM, the decay of a modulus dilutes baryon asymmetry existing from any previous era and has to be regenerated. This turns out to be quite natural. The decay of the modulus provides the non-equilibrium condition required for baryogenesis and a typical value for the dilution factor\(^4\) is \( Y \sim 10^{-8} \). If the modulus decays to a field that has CP- and B- violating couplings to SM fields, baryogenesis can occur through usual interference between tree and loop-level diagrams. These interference terms are usually suppressed by factors like \((1/4\pi)\), so that a net factor of \( \sim 10^{-10} \) for the baryon asymmetry can be obtained with \( \mathcal{O}(1) \) Yukawa couplings [261, 336]. For a more detailed discussion on baryogenesis, see Sec. 3.2.

2.5. Post-BBN modifications (Authors: T. Karwal & T. L. Smith)

In this section, we discuss causes of nonstandard, post-BBN expansion history. We outline examples of modifications to the properties of the standard ΛCDM components, and scenarios which introduce new components altogether.

These two categories can be further refined into modifications to the background and/or perturbative evolution. The conservation of stress-energy dictates how all minimally coupled components evolve in both the (homogeneous and isotropic) background as well as at linear order in their inhomogeneous perturbations (see e.g. Ref. [342]):

\[
\dot{\rho} = -3\frac{\dot{a}}{a} \rho (1+w),
\]
\[
\ddot{\delta} = -(1+w)(\theta + 3\dot{\phi}) - 3\frac{\dot{a}}{a} \left( c_s^2 - w \right) \delta,
\]
\[
\dot{\theta} = -\frac{\dot{a}}{a}(1 - 3w)\theta - \frac{\dot{w}}{1+w}\theta + \frac{c_s^2}{1+w} k^2 \delta - k^2 \Sigma + k^2 \psi,
\]

where dots indicate derivatives with respect to conformal time, \( \dot{\rho} \) is the fluid’s homogeneous density, \( w \equiv \dot{P}/\rho \) is its equation of state parameter, \( \delta \equiv \delta \rho/\rho \) is its density contrast, \( \theta \) is the divergence of its velocity perturbation, \( c_s^2 \equiv \delta P/\delta \rho \) is its sound speed, \( \phi \) and \( \psi \) are the temporal and spatial gravitational metrics in conformal Newtonian gauge, and \( \Sigma \) is the fluid’s anisotropic stress.

The simplest modifications to the ΛCDM expansion history can be characterised as changes to \( w \) in Eq. (20). Models which do this can generically be characterised by the way in which the equation of state parameter, \( w \), changes with time. For example, quintessence models which promote the cosmological “constant” to a dynamical quantity predict some time variation in the dark energy equation of state, \( w_{de} \), affecting the late-time expansion history relative to “\( \Lambda \)”-CDM. Particular quintessence models have the property that \( w_{de} \) tracks the matter equation of state parameter, \( w_m \) [343–346]. These “tracker” models are particularly interesting as potential solutions to the coincidence problem of dark

\(^4\) The dilution factor is defined as \( Y \equiv n_\sigma/s = 3T_{reh}/(4m_\sigma) \), where the last expression can be obtained directly from the Boltzmann equation for \( n_\sigma \).
energy. Such tracker dark energies can arise from various mechanisms including scalar fields with particular potentials [344], quintessence and k-essence models [347–349] and alternative models of gravity [346]. Models which propose additional forms of energy density, beyond what is contained in ΛCDM, can also modify the post-BBN expansion history. A universe which contains a collection of scalar fields may undergo periodic epochs of anomalous expansion (see e.g. Refs. [350, 351]). Some models of minimally coupled “early” dark energy (EDE) envision a scalar field which is fixed at some point in its potential by Hubble friction which becomes dynamical when the Hubble parameter drops below some critical value around the matter-radiation equality [352, 353]. The fractional contribution of this field to the total energy density briefly increases and then decreases, as long as it dilutes faster than matter. Other models propose a nonminimal coupling to other components (such as the neutrino sector) which then produces a “kick” to the scalar field, again leading to a brief increase in the total energy density of the Universe [354]. Due to its success at providing a compelling increase to the Hubble tension [355], numerous particle models have now been proposed to comprise of such EDEs, including axion-like [353, 356–358] and power-law scalar fields [359] with nonstandard kinetic terms [360], and first-order phase transitions in the dark sector [361].

Several models of nonstandard expansion also arise due to interactions between the different components that fill the Universe. Conservation of stress-energy in the presence of exchange of energy-momentum between different components (X and Y) can be written as (see e.g. Ref. [362])

\[ \nabla_\mu T^\mu_{XY} = Q_Y, \quad \nabla_\mu T^\mu_{YX} = -Q_Y, \]

(23)

where \( Q_Y \) is the energy-momentum flux exchanged between the two components. The form of \( Q_Y \) depends on the specific physics of the coupling, but a natural choice is to write the energy-momentum transfer as a linear combination of the four-velocities [363]

\[ Q_Y = \zeta_X u_{X,Y} + \zeta_Y u_{Y,X}, \]

(24)

where \( u_{(X,Y)X} \) is the four-velocity of the components. Models with interactions of this form include interactions between dark matter and dark energy [364], developed in order to address the coincidence problem [365, 366]. A similar inter-species interaction is DM decaying into dark radiation [367, 368], invoked to explain the secular and Hubble tensions. While there are stringent bounds on all DM behaving this way, a fraction of DM annihilating into dark radiation is permitted by cosmological data. Depending on the rate \( \Gamma \) of decay of the DM, when the expansion rate \( H \sim \Gamma \), all the interacting DM decays into dark radiation over one cosmic time scale, which rapidly dilutes away. If this event occurs during MD, the expansion of the Universe is suddenly slowed by a factor roughly proportional to the square-root of the fraction of decaying DM, and then follows the usual dilution of matter density again. Close to matter-radiation equality or matter-dark-energy equality, such an event can shift the redshift of equality. At other times, when matter is subdominant, the effects of decaying DM on the background evolution of the Universe are minimised, with the principal effects being confined to the evolution of perturbations.

Given the form of the energy-momentum exchange rate in Eq. (24), at the background level we have

\[ \dot{\rho}_X + 3H \rho_X (1 + w_X) = Q_0, \]
\[ \dot{\rho}_Y + 3H \rho_Y (1 + w_Y) = -Q_0, \]

(25)

which generalize Eq. (11). The impact of this interaction on the expansion rate can be more simply understood once we realize that the sum of the continuity equations for the interacting components is, itself, conserved. Therefore, at the background level, the two components behave as a single, effective, component with an effective equation of state parameter \( w_{\text{eff}} \equiv (\dot{\rho}_X u_X + \dot{\rho}_Y u_Y)/(\dot{\rho}_X + \dot{\rho}_Y) \). In the noninteracting case, \( Q_0 = 0 \), for constant equation of state parameters, we know that \( \rho_{X,Y} \propto a^{-3(1+w_{X,Y})} \) so that, taking \( w_X > w_Y, w_{\text{eff}} \) monotonically transitions from \( w_X \) to \( w_Y \) as the Universe expands. In the interacting case, \( Q_0 \neq 0 \), as energy density flows from one component to the other, \( w_{\text{eff}} \) will deviate from a monotonic evolution, leading to a change in the expansion history. Note that this effective fluid description is capable of producing a component with \( w_{\text{eff}} < -1 \) even if \( w_X, w_Y \geq -1 \) [369]. A similar effect is seen when working out the dynamics in some modified gravity theories. For example, we can write any scalar-tensor gravity theory in the Einstein frame, in which the modified expansion is due to a nonminimal coupling between the scalar field and other components [370]. In addition to this, an agnostic treatment of the microscopic dynamics of dark energy makes it indistinguishable from modified gravity [371]. This effective treatment of modified gravity and nonstandard dark energy interactions is the main focus of the “parameterized post-Friedmann” formalism [372].

Finally, DM-baryon interactions [373–375] have been recently considered as a potential explanation of the anomalous measurements from EDGES [37]. In these models the DM-baryon cross-section behaves as \( \sigma = \sigma_0 v^n \), where \( v \) is the relative velocity between the scattering particles. This velocity dependence may arise in several natural interactions, such as electric- or magnetic-dipole interactions through light mediators \( (n = -2) \) or Coulomb-like interactions or Yukawa interactions through light mediators \( (n = -4) \).

It is important to note that when the background equations are modified, in order to maintain energy-momentum conservation at the linear level, the perturbation equations must also be modified. However, there are interactions that preserve the background equations and just modify the evolution of the linear perturbations. Thomson scattering between photons and baryons and any self-interactions are examples of such mechanisms. In the remainder of this section we will consider the effects of scattering so that the number of each constituent remains unchanged.
For a self-interacting relativistic component [376, 377], the linear conservation equations [Eqs. (21) and (22)], remain unchanged. The modifications enter through the evolution of the component’s anisotropic stress, Σ. The evolution of Σ is governed by the Boltzmann equation which, in turn, is sourced by a collision term. In many cases this collision term can be written as $C[F] = \mathcal{F}/\tau$, where $\mathcal{F}$ is the component’s distribution function and $\tau$ is the rate at which the component’s opacity changes (see e.g. Ref. [378]). In the absence of this collision term, the equations of motion for the anisotropic stress Σ cause a decrease over time of the density contrast and velocity perturbation (i.e. “free-streaming”); self-interactions disrupt this and lead to enhanced clustering for that component. On the other hand, self-interacting DM [379, 380] has been considered to alleviate a variety of small-scale structure “problems”. However, the typical interaction cross-sections that are considered are only active in collapsed, nonlinear, structures and hence have little effect on the linear physics probed by the CMB and large-scale structure observations.

Models which excite initial conditions beyond the standard adiabatic initial conditions also modify the late-time evolution of perturbations. Any model of inflation that involves more than one dynamically important scalar field, such as the curvaton scenario [381] or axion DM [382], produce nonadiabatic initial conditions, which will be discussed in more detail in Sec. 3.4 (see also Ref. [383]). As a specific example, we mention compensated isocurvature perturbations which balance an overdensity of baryons with an underdensity of cold DM and vice-versa (see e.g. Refs. [384, 385]). This scenario leaves the local matter density perturbations unchanged but affects the plasma sound speed $c_s^2$, leading to a smoothing of primordial perturbations. This has been suggested as a solution to the $A_{\text{obs}}$ anomaly [386, 387] and may partially alleviate the Hubble tension through large-scale fluctuations in $\Omega_b$ [388].

Finally, models which promote “constants” of nature to dynamical fields may also cause nonstandard cosmological evolution. Scalar-tensor theories of gravity, mentioned before, are an example of this, leading to a dynamical gravitational coupling, $G(\bar{x},t)$. Other intriguing examples are models of a dynamical fine-structure constant [389, 390]. These models predict a time and spatial variation in the strength of electromagnetic interactions (such as Thomson scattering) which has a unique impact on the structure of the observed CMB anisotropies (see e.g. Ref. [391, 392]).

### 3. Consequences of Nonstandard Expansion Phases

In this section, we will discuss several consequences of nonstandard expansion eras, paying particular attention to models of dark matter, early Universe phase transitions and baryogenesis, models of inflation, and generation of the primordial curvature perturbations, as well as microhalos and PBHs.

#### 3.1. Consequences for dark matter (Authors: A. Berlin, D. Hooper & G. Krnjaic)

If the early Universe was not consistently dominated by radiation as is typically assumed, the phenomenology of DM production and detection could be substantially altered from standard expectations. In particular, as discussed in Sec. 2.3, if the visible sector (which contains the SM) is supplemented by a decoupled hidden (or dark) sector (which contains the DM), these two sectors can be produced independently during post-inflation reheating and maintain separate temperatures throughout cosmological evolution. Within the hidden sector, the DM could annihilate to lighter hidden sector states which ultimately decay into SM particles. If these unstable states are sufficiently heavy and long-lived, they will come to dominate the energy density of the Universe before decaying. The lightest hidden sector state will automatically be long-lived, since its width relies on a coupling that is too small to sustain thermal equilibrium between the two sectors. This picture is relatively universal, and can be found within any model in which the DM freezes out through annihilations in a heavy and highly decoupled hidden sector that is populated after inflation. In contrast to scenarios in which an additional out-of-equilibrium decay is invoked solely to dilute the initial cosmological abundances of various species, dilutions of this kind are an inevitable consequence of thermal decoupling.

Although this class of scenarios can be realized within the context of a wide range of hidden sector models, we will consider a simple vector portal model as an illustrative example. For our DM candidate, we introduce a stable out-of-equilibrium decay is invoked solely to dilute the initial cosmological abundances of various species, dilutions of and highly decoupled hidden sector that is populated after in

These models predict a time and spatial variation in the strength of electromagnetic interactions (such as Thomson scattering) which has a unique impact on the structure of the observed CMB anisotropies. However, the typical interaction cross-sections that are considered are only active in collapsed, nonlinear, structures and hence have little effect on the linear physics probed by the CMB and large-scale structure observations.

Long lifetimes are straightforwardly realized if the decaying particle is the lightest hidden sector state. In fact, if the hidden and visible sectors are decoupled, the lightest hidden sector state will automatically be long-lived, since its width relies on a coupling that is too small to sustain thermal equilibrium between the two sectors. This picture is relatively universal, and can be found within any model in which the DM freezes out through annihilations in a heavy and highly decoupled hidden sector that is populated after inflation. In contrast to scenarios in which an additional out-of-equilibrium decay is invoked solely to dilute the initial cosmological abundances of various species, dilutions of this kind are an inevitable consequence of thermal decoupling.

Finally, models which promote “constants” of nature to dynamical fields may also cause nonstandard cosmological evolution. Scalar-tensor theories of gravity, mentioned before, are an example of this, leading to a dynamical gravitational coupling, $G(\bar{x},t)$. Other intriguing examples are models of a dynamical fine-structure constant. These models predict a time and spatial variation in the strength of electromagnetic interactions (such as Thomson scattering) which has a unique impact on the structure of the observed CMB anisotropies.
for $m_X \gg m_{Z'}$, this cross-section is given by $\sigma v \approx \pi \alpha_X^2 / m_X^2$, where $\alpha_X \equiv g_{0,DM}^2 / 4\pi$.

As an initial condition, we take the hidden and visible sectors to be described by separate thermal distributions, with temperatures of $T_h$ and $T$, respectively. The ratio of these temperatures, $\xi_{\text{inf}} \equiv (T_h/T)_{\text{inf}}$, is determined by the physics of inflation, including the sectors’ respective contributions to the inflaton [423, 424]. From entropy conservation in each sector, we can calculate the time evolution of $\xi$ (prior to the decays of $Z'$)

$$\frac{s_h}{s} = \frac{g_{*,h}}{g_{*,v}} \xi^3 = \text{constant} \rightarrow \xi = \xi_{\text{inf}} \left( \frac{g_{*,h}}{g_{*,v}} \right)^{1/3} \left( \frac{g_{*,s}}{g_{*,s,\text{inf}}} \right)^{1/3}, \tag{27}$$

where $s$ and $s_h$ are the hidden and visible sector entropy densities, $g_{*,h}$ and $g_{*,v}$ are the numbers of effective relativistic entropy degrees of freedom in the visible and hidden sectors. If the SM temperature is well above the electroweak scale, $g_{*,v} \approx g_{*,v,\text{inf}}$. As the temperature of the hidden sector falls below $m_X$, $g_{*,h}$ decreases from $g_{Z'} + (7/8)g_{X}$ to $g_{Z'}$, bringing $\xi$ from $\xi_{\text{inf}}$ to $(13/6)^{1/3} \xi_{\text{inf}} \approx 1.3 \xi_{\text{inf}}$, for $m_{Z'} \ll m_X$.

As the Universe expands, $X$ will eventually freeze-out of chemical equilibrium, yielding a nonnegligible relic abundance. The evolution of the number density of $X$ (plus $X$), $n_X$, is described by the Boltzmann equation

$$\dot{n}_X + 3Hn_X = -\frac{1}{2} \langle \sigma v \rangle \left( n_X^2 - \frac{n_{Z',\text{eq}}^2}{n_{X,\text{eq}}^2} \right), \tag{28}$$

where $\langle \sigma v \rangle$ is the thermally averaged cross-section for the process $X \bar{X} \rightarrow X' Z'$, and $H = [8\pi (\rho_R + \rho_H)/3 M_P^2]^{1/2}$ describes the expansion rate of the Universe in terms of the energy densities in the visible and hidden sectors.

In the case that $n_{Z'}$ remains close to its equilibrium value during the freeze-out of $X$, the Boltzmann equation can be solved semi-analytically. In this case, the thermal relic abundance of $X$ (plus $X$) is given by

$$\Omega_X h^2 \approx 1.7 \times 10^{-10} \frac{x_f \sqrt{g_{\text{eff}}}}{g_{*,s}} \left( \frac{\sigma v_{X \bar{X} \rightarrow Z' Z}}{\text{GeV}^{-2}} \right)^{-1} \approx 1.6 \times 10^4 \left( \frac{x_f}{30} \right) \left( \frac{0.1}{\alpha_X} \right)^2 m_X^{\frac{1}{2}} \left( \frac{\sqrt{g_{\text{eff}}}/g_{*,s}}{0.1} \right), \tag{29}$$

where $g_{\text{eff}} \equiv g_{*,s} + g_{*,s}^{h} \xi^4$ at freeze-out. $x_f$, which is defined as the mass of $X$ divided by the SM temperature at freeze-out, is found to be $\sim 20 \times \xi$ over a wide range of parameters. From Eq. (29), it is clear that a PeV-scale DM candidate with perturbative couplings will initially freeze-out with an abundance that exceeds the observed DM density ($\Omega_X h^2 \gg \Omega_{DM} h^2 \approx 0.12$). It has long been appreciated, however, that this conclusion can be circumvented if the Universe departed from the standard RD picture after DM freeze-out [214, 262, 279, 324, 395–400, 425–427]. More specifically, as the Universe expands, the remaining $Z$’s will become nonrelativistic and quickly come to dominate the energy density of the Universe when $\rho_{Z'} = 0.0074 g_{*,s,\text{inf}}^{3} m_{Z'} T_{\text{dom}}^{3} > (\pi^2/30) g_{*,s,\text{inf}}^{3} T_{\text{dom}}^{4}$, which occurs at a visible sector temperature

$$T_{\text{dom}} \sim 1 \text{ TeV} \times \xi_{\text{inf}}^{3} \left( \frac{m_{Z'}}{50 \text{ TeV}} \right). \tag{30}$$

This expression is valid so long as the $Z'$ population departs from chemical equilibrium while relativistic. When the $Z$’s ultimately decay (see Fig. 4), they will deposit energy and entropy into the visible sector, potentially diluting the DM abundance to acceptable levels. In the left frame of Fig. 5, we show the evolution of the energy densities in the visible and hidden sectors, for a representative choice of parameters in this model.

To calculate the impact of this effect, we integrate over the $Z'$ decay rate to find the factor by which the relics will
be diluted [15]
\[
\frac{s_f}{s_i} \approx 680 \times \left( \frac{10^{-10}}{\epsilon} \right) \left( \frac{m_{Z'}}{100 \text{ TeV}} \right)^{1/2} \left( \frac{g^*_{s,s}}{100} \right)^{1/4} \xi_{\inf}^3,
\]
where \((g^*_{s,s})\) denotes the time-averaged value over the period of decay. Combining this with Eq. (29), we find that the final DM relic abundance is given by
\[
\Omega_X h^2 \approx \frac{0.12}{\xi_{\inf}^3} \left( \frac{\epsilon}{10^{-13}} \right) \left( \frac{0.045}{\alpha_X} \right)^2 \left( \frac{m_X}{\text{PeV}} \right)^2 \left( \frac{100 \text{ TeV}}{m_{Z'}} \right)^{1/2} \left( \frac{x_f}{30} \right) \left( \frac{\sqrt{y_{\text{eff}}}/g^*_{s,s}}{0.1} \right) \left( \frac{100}{g^*_{s,s}} \right)^{1/4} \left( \frac{1/3}{\xi_{\inf}^3} \right)^{1/4}.
\]

In the right panel of Fig. 5 we plot some of the phenomenological features of this model as a function of the DM mass and the degree of kinetic mixing between the hidden sector and SM hypercharge. The black contours denote the regions where the DM density is equal to the measured cosmological abundance, for three values of the hidden sector interaction strength, \(\alpha_X\). Below the brown region, \(Z'\) decays deposit significant entropy into the visible sector, reducing the final X abundance. To assure consistency with BBN (see Sec. 4.1 for details), we require that the temperature of the Universe exceeds 10 MeV after the decays of the \(Z'\) population, resulting in the following constraint:
\[
\epsilon \geq 2 \times 10^{-13} \times (100 \text{ TeV}/m_{Z'})^{1/2} (g^*_{s,s}/10)^{1/4}.
\]

Viable phenomenology can be obtained within this class of models for a wide range of portal couplings, spanning many orders of magnitude. Depending on the degree of kinetic mixing, the hidden and visible sectors may have been entirely decoupled from one another, or kept in kinetic equilibrium through interactions of the type \(\gamma f \leftrightarrow Z'f\). Quantitatively, we find that the rate for these processes exceed that of Hubble expansion if:
\[
\epsilon \geq 10^{-7} \times (T/10 \text{ GeV})^{1/2}
\]
(known as the orange region in the right panel of Fig. 5). Thus, for smaller values of \(\epsilon\), the hidden sector will not reach equilibrium with the visible sector and will remain decoupled. Furthermore, in the yellow regions of the right panel of Fig. 5, the \(Z'\) population decays prior to the freeze-out of \(X\).

Many hidden sector DM models are difficult to constrain or otherwise test. In particular, the feeble couplings between the hidden and SM sectors makes the prospects for DM searches at colliders and at direct detection experiments quite bleak (see, however, Ref. [428]). This is less true for indirect searches. In particular, during the EMDE that is many orders of magnitude. Depending on the degree of kinetic mixing, the hidden and visible sectors may have could form during the RD era. In Ref. [341], a calculation was performed of the evolution of density perturbations within the context of such hidden sector models, using a series of \(N\)-body simulations to determine the outcome of nonlinear collapse during a RD era. The resulting microhalos were found to be extremely dense, leading to very high rates of DM annihilation and to large indirect detection signals that resemble those ordinarily predicted for decaying.
DM. Existing measurements of the high-latitude gamma-ray background rule out a wide range of parameter space within this class of models. The scenarios that are most difficult to constrain are those that feature a very long EMDE; if microhalos form prior to the decay of the unstable hidden sector matter, the destruction of these microhalos effectively heats the DM, suppressing the later formation of microhalos. This gravitational heating is discussed in more detail in Sec. 3.5.

3.2. Phase transitions and baryon asymmetry (Authors: R. Allahverdi & M. Lewicki)

3.2.1. Phase transition dynamics

Let us begin with discussion of the modification of dynamics of a phase transition due to modified expansion history. While the modification itself can seem to be rather drastic, phase transitions prove to be largely insensitive. This is simply because the dynamics of a phase transition are largely dictated by the dynamics of fields undergoing the transition. For cosmological applications, in the most common case of a thermally driven transition of a scalar field, decay rate per unit volume is given by [215–218]

$$\Gamma(T) \approx T^4 \exp \left( -\frac{S_3(T)}{T} \right), \quad S_3 = 4\pi \int d^2 r \left[ \frac{1}{2} \left( \frac{d\phi}{dr} \right)^2 + V(\phi, T) \right],$$

where the action $S_3(T)$ depends on the potential of the field in question including thermal corrections. Expansion history comes into play when we find the temperature at which bubbles begin nucleating defined by

$$\int_0^{t_n} \frac{dt}{H(t)^3} = \int_{T_n}^{\infty} \frac{dT}{T} \frac{\Gamma(T)}{H(T)^3} = 1,$$

where we assumed adiabatic expansion, that is $dt = -d\tau/T(H(T))$. Given the exponential dependence of the decay rate we can see that modified expansion history will impact the resulting temperature only logarithmically (see e.g. Ref. [429]). Thus, phase transitions are rather insensitive to expansion history and their course is instead linked to the (subdominant) radiation sector the fields in question are coupled to.

With that said, it is of course true that possible observable consequences of such a transition can still be impacted a lot more severely due to the modified expansion. We will discuss these modification in terms of electroweak baryogenesis in the remainder of this section and the gravitational wave background in Sec. 4.4.

3.2.2. Electroweak baryogenesis

Let us now turn to electroweak baryogenesis [430–433] in which the observed baryon asymmetry of the Universe is a result of a strong electroweak phase transition. This requires a modification of the SM to make the electroweak phase transition first order and introduce extra sources of CP-violation. Our goal, however, is not to discuss any specific modification of SM aimed at satisfying these conditions but rather to discuss how they can be modified in nonstandard expansion histories.

We will focus on the main constraint typically put on the transition itself, that is the vacuum expectation value of the field over the temperature has to be large enough $v/T \gtrsim 1$. This comes from the requirement that the electroweak sphalerons providing required baryon number violation are suppressed after the transition to avoid washing out the asymmetry as the system goes back to thermal equilibrium. To be more specific, the sphaleron processes in question are suppressed after the transition to avoid washing out the baryon asymmetry as the system goes back to thermal equilibrium. To be more specific, the sphaleron processes in question are suppressed after the transition to avoid washing out the baryon asymmetry (i.e. $v/T \gtrsim 1$) when the expansion is driven by SM radiation $H(T) = H\text{R}(T) \propto T^2$. The modification of the required $v/T$ is shown in Fig. 6 assuming the transition occurs at the typical temperature for electroweak symmetry breaking $T \approx 100$ GeV. While lowering the value of $v/T$ required to avoid wash-out by sphalerons might not seem like a huge modification, it has a significant effect on the parameter space of specific models [439, 440].

As discussed in Sec. 2, a particularly well-motivated example of a nonstandard thermal history involves an EMDE. The EMDE ends upon completion of the decay and the Universe enters a RD phase with temperature $T_{\text{reh}}$. Electroweak baryogenesis in such a scenario with $1 \text{ GeV} \lesssim T_{\text{reh}} \ll 100 \text{ GeV}$ has been studied in Ref. [260]. There it was found that a faster expansion rate at the same temperature results in a bound $v/T \gtrsim 0.77$ (0.92) for $T_{\text{reh}} = 1$ (10) GeV. Note that a modified expansion rate affects the out-of-equilibrium condition necessary for generation of baryon asymmetry (i.e.
the third Sakharov condition \[441\]), as well as the efficiency of wash-out processes that could erase the asymmetry, in any baryogenesis scenario.

### 3.2.3. Constraints on baryogenesis from dilution by reheating

In general, a nonstandard thermal history can also affect baryogenesis by diluting the generated baryon asymmetry. This is due to entropy release during transition to the final RD phase. This can be readily seen by noting that the energy density of a component with equation of state $w < 1/3$ is redshifted more slowly than that of radiation. Thus its dominance keeps growing and there can be no transition to RD era unless it decays to SM radiation. This reheating process is accompanied by (a typically significant) increase in the entropy of the Universe.

To demonstrate the effect of reheating on baryon asymmetry, we consider a nonstandard thermal history that involves a phase with equation of state $w < 1/3$ that is preceded and followed by RD era. Let us denote the Hubble rate at the onset of the nonstandard expansion phase or generation of the baryon asymmetry (whichever happened later) by $H$ and at the time of return to RD era by $H_{\text{reh}}$. The number density of any quantity produced at $H > H_{\text{reh}}$ is redshifted by a factor $(a_{\text{reh}}/a)^3 = (H/H_{\text{reh}})^{2/(1+w)}$ between $H$ and $H_{\text{reh}}$. Applying this to the normalized baryon asymmetry $n_B/s$, where $n_B$ is the number density of baryon asymmetry, the maximum dilution factor during a nonstandard expansion phase with $w < 1/3$ is found to be

$$d_{\text{max}} \sim \left(\frac{H}{H_{\text{reh}}}ight)^{(1-3w)/(1+w)}.$$

Fig. 6 shows the dilution factor in the case $w = 0$ as a function the Hubble rate normalised to the Hubble rate at reheating $H_{\text{reh}}$.

To put things in perspective, let us consider an EMDE ($w = 0$) driven by some string modulus $\phi$ with mass $m_\phi$. We typically have $H \sim m_\phi$ at the onset of the EMDE and $H_{\text{reh}} \sim m_\phi^3/M_P^2$ in this case (see Sec. 2.4), which results in $d_{\text{max}} = O(M_P/m_\phi)$. Considering that $m_\phi \gtrsim O(100 \text{ GeV})$ is needed in order to to reheat the Universe before the onset of BBN (i.e. $T_{\text{reh}} \gtrsim O(1) \text{ MeV}$), we find $d_{\text{max}} \lesssim 10^{-3} - 10^{-4}$. Such a large dilution can render any previously generated baryon asymmetry totally insignificant.

Therefore, to obtain the observed value $(n_B/s) \approx 0.9 \times 10^{-10}$, we are left with two general options. First, to generate an initially large baryon asymmetry that matches the correct value after dilution. Second, to generate the correct baryon asymmetry after the transition to the final RD phase thereby avoiding dilution by reheating. Let us discuss these possibilities and specific baryogenesis mechanisms that can work in each case in some detail.

Affleck-Dine (AD) baryogenesis \[260, 442\] (for reviews, see Refs. \[443-445\]) is a mechanism that can naturally fit the first option. This scenario utilizes flat directions in the scalar potential of supersymmetric extensions of the SM. AD baryogenesis is known to be able to generate large values of baryon asymmetry, even $O(1)$, in a standard thermal history. An EMDE (driven, for example, by string theory moduli) can be therefore helpful in this regard by regulating the baryon asymmetry and bringing it down to the observed value \[337, 446-448\]. A viable string embedding of AD baryogenesis has been discussed in Ref. \[449\] where the volume modulus in type IIB string models drives an EMDE phase resulting in the observed baryon asymmetry for natural values of the underlying parameters.

The second route is to generate baryon asymmetry upon completion of transition to a RD Universe when the Hubble rate is $H_{\text{reh}}$. Many scenarios with a nonstandard thermal history result in $T_{\text{reh}} \ll 100 \text{ GeV}$. In particular, in cases where an EMDE is responsible for nonthermal production of weak-scale DM, we may have $T_{\text{reh}} \lesssim O(1 \text{ GeV})$. This
severely constrains viable baryogenesis mechanisms that can work. For example, it rules out leptogenesis [49] (for reviews, see Refs. [450, 451]) as sphalerons are out-of-equilibrium at these temperatures and hence unable to convert the generated lepton asymmetry to baryon asymmetry. Thermal baryogenesis is essentially ruled out as a possibility too because it will require C-, CP-, and B-violating interactions at such low scales.

Therefore, nonthermal post-sphaleron baryogenesis seems to be the available possibility for generating the observed baryon asymmetry at late times. An explicit model is discussed in Ref. [261] in the context of EMDE driven by some string modulus $\sigma$. This model includes new particles $N$ with a mass $m_N \gg T_{\text{reh}}$ that have C-, CP-, and B-violating interactions with the SM particles. $N$ quanta are produced in $\sigma$ decay, and their subsequent decay to the SM particles generates a baryon asymmetry given by

$$\frac{n_B}{s} = Y_\sigma \text{ Br}_N \epsilon, \quad Y_\sigma \equiv \frac{3T_{\text{reh}}}{4m_\sigma}. \quad (37)$$

Here $Y_\sigma$ is the yield from $\sigma$ decay, Br$_N$ is the branching fraction for production of $N$ from $\sigma$ decay and $\epsilon$ is the asymmetry generated per $N$ decay. One point to keep in mind is that $Y_\sigma$ is typically very small. For example, a modulus with mass $m_\sigma \sim 100 - 1000$ TeV gives $Y_\sigma \lesssim \mathcal{O}(10^{-6})$. Therefore, Br$_N$ cannot be too small in order to generate the correct value of $n_B/s$. This implies an $\epsilon$ usually larger than that in thermal scenarios.

This nonthermal scenario may also address the baryon-DM coincidence problem (for a review, see Ref. [452]) if $\sigma$ decay is the common origin of baryogenesis and DM [336, 453]. In addition, if $N$ particles are not much heavier than $\mathcal{O}(1)$ TeV, they could be produced at the LHC. The $B$-violating interactions of $N$ can also lead to observable low-energy phenomena like neutron-antineutron oscillations. Such a TeV scale baryogenesis scenario may therefore be tested by complementary probes at the cosmology, high energy, and low energy frontiers [454].

So far, our quantitative discussion of the effects of a nonstandard thermal history on baryogenesis has mainly focused on EMDE as an important and well-motivated case. However, the main considerations regarding the out-of-equilibrium condition to generate and preserve baryon asymmetry, and the entropy release during transition to the final RD phase are valid for any scenario with a modified expansion rate. Here, we mention in passing a number of studies on baryogenesis in more general nonstandard thermal histories. TeV scale leptogenesis in nonstandard thermal histories that involve a) an EMDE, and b) a phase with equation of state $w > 1/3$ has been discussed in Ref. [410]. An important example of a scenario with $w > 1/3$ is kination where $w = 1$. Ref. [455] considers AD baryogenesis in the context of a model that involves a post-inflationary kination phase. The effect of kination on the gravitational leptogenesis scenario has been investigated in Ref. [456]. Gravitational leptogenesis for a general equation of state during inflationary reheating has been studied in Ref. [57]. It has been shown in Ref. [457] that nonstandard cosmologies from scalar-tensor theories can open new paths to low-scale leptogenesis and thereby allow direct experimental tests of the scenario. Finally, it is possible to have more complicated situations that involve various sectors with different temperatures. A specific example of such a scenario is given in Ref. [263] that considers thermal leptogenesis in the presence of a dark sector whose temperature is higher than that of SM radiation.

3.3. Consequences for models of inflation (Authors: K. Freese & T. Tenkanen)

Cosmic inflation can be regarded as a successful paradigm for two reasons. First, if inflation lasted long enough so that during it the Universe expanded at least by a factor$^5$ of $\sim e^{60} \sim 10^{26}$, inflation provides a plausible explanation for the classic horizon, monopole, and flatness problems (see e.g. Ref. [458]). Second, inflation provides a natural way to generate the initial metric perturbations and explain their super-horizon correlations at the time of CMB decoupling. These correlations are usually described in terms of the scalar curvature power spectrum, which at least at the largest physical distance scales is measured to have a power-law form [13]

$$P_\zeta(k) = A \left(\frac{k}{k_s}\right)^{n_s - 1}, \quad (38)$$

to a high precision. Here $\zeta$ denotes the curvature perturbation and $k$ the wavenumber in a Fourier decomposition. The spectrum (38) has the observed amplitude $A \approx 2.1 \times 10^{-9}$ and spectral tilt $n_s \approx 0.965$ at the pivot scale $k_s = 0.05 \text{ Mpc}^{-1}$ [13]. The data are consistent with no running, $d\eta_s/d\ln k$, or running of the running of the spectral tilt, $d^2\eta_s/d(\ln k)^2$ [13], and therefore we will neglect them here for simplicity. While in principle solving the classic horizon, monopole, and flatness problems only requires exponential expansion at early times, any successful model of inflation also has to predict the measured values for observables, or be within the limits on them. In addition to depending on the underlying particle physics scenario, these predictions usually depend also on the post-inflationary expansion history, which is the topic of this section.

In the following, we will assume, for simplicity, that inflation was driven by a single scalar field $\phi$, which at early times dominated the total energy density of the Universe and drove inflation by rolling down in its potential $V$. In the slow-roll approximation, $|\dot{\phi}| \ll 3H|\phi|, \dot{\phi}^2 \ll V$, where $H \equiv \dot{a}/a$ is the Hubble parameter and the dot denotes

$^5$ Assuming high scale inflation and standard expansion history. In nonstandard cosmologies, this number can range from $\sim e^{18}$ to $e^{68}$ [20–22].
derivative with respect to cosmic time, the inflationary dynamics can be characterized by the slow-roll parameters

\[ \epsilon \equiv \frac{M_p^2}{16\pi} \left( \frac{V'}{V} \right)^2, \quad \eta \equiv \frac{M_p^2}{8\pi} \frac{V''}{V}, \]

where the prime denotes derivative with respect to \( \phi \). In slow-roll \( \epsilon, |\eta| \ll 1 \), and inflation ends when \( \epsilon = 1 \). In this case, then, the amplitude of the curvature power spectrum can be expressed in terms of the slow-roll parameters as \[ \mathcal{A} = \frac{1}{24\pi^2 M_p^2} \frac{V(\phi_*)}{\epsilon(\phi_*)}, \]

where \( \phi_* \) denotes the field value at the time when the scale \( k_* \) exited the horizon. The leading order expression for the spectral tilt is

\[ n_s - 1 \equiv \frac{d \ln P_{\mathcal{R}}(k)}{d \ln k} \bigg|_{k=k_*} \simeq -6\epsilon(\phi_*) + 2\eta(\phi_*) . \]

In addition to generating scalar curvature perturbations, fluctuations of the inflaton field also generate tensor perturbations, i.e. primordial gravitational waves. The observational constraints are usually expressed in terms of the tensor-to-scalar ratio

\[ r = \frac{P_T}{P_s} \simeq 16\epsilon, \]

where \( P_T \) is the tensor power spectrum (see e.g. Ref. [458]) and the last expression applies at the leading order in slow-roll parameters. As primordial tensor perturbations have not been detected, observations of the CMB place an upper limit on tensor-to-scalar ratio, \( r < 0.06 \) at the pivot scale \( k_* = 0.05 \text{Mpc}^{-1} \) \[460\].

As can be seen by substituting the slow-roll parameters in Eq. (39) into Eqs. (40), (41), and (42), the inflationary observables depend solely on the shape of the inflaton potential. Constraints on the observables therefore limit different scalar field potentials, i.e. different models, at the energy scale where inflation occurred. However, judging whether a given model complies with observations is not entirely clear-cut. A given potential may predict wrong values for observables at some energy scale but get them exactly right if the distance scales where measurements are made exited the horizon when the inflaton was rolling in some other part of the potential where its shape was different and which therefore generated different values for observables. The time when a given scale must have exited the horizon is determined by the overall energy scale of inflation and the post-inflationary evolution of scales. Therefore, details of the entire post-inflationary expansion history and, in particular, any nonstandard period of expansion will affect inflationary model selection, as we will demonstrate below.

It is convenient to characterize the observables in terms of the number of e-folds between the horizon exit of the pivot scale and the end of inflation, \( N \equiv \ln(a_{\text{end}}/a_k) \) where \( a_{\text{end}} \) is the value of the scale factor at the end of inflation (when the scale \( k \) exited the horizon), which – together with the parameters that characterize the post-inflationary evolution – relate the pivot scale to the present Hubble scale \( H_0 \) as

\[ \frac{k}{a_0H_0} = \frac{a_kH_k}{a_0H_0} = e^{-N} \frac{a_{\text{end}}}{a_{\text{end}}} \frac{a_{\text{RD}}}{a_{\text{RD}}} \frac{H_k}{H_0} , \]

where \( H_k \) is the Hubble scale at \( a_k \), \( a_{\text{RD}} \) is the scale factor at the time when the nonstandard period ended and the Universe became RD, and \( a_0 \) is the scale factor at present. Here we have assumed that there is exactly one nonstandard phase between the end of inflation and the subsequent RD period. Note that the nonstandard expansion period could have been caused by, for example, a kinetic energy-dominated kination phase \[255,434\] or another period of inflation (see Sec. 2.2) instead of conventional reheating discussed in Sec. 2.1. Including another nonstandard expansion phase(s) that punctuated the subsequent RD era(s) is straightforward \[20\] but here we leave them out for simplicity. Assuming that the nonstandard phase can be characterized with a single effective equation of state parameter \( w \), so that the total energy density scaled as \( \rho \propto a^{-3(1+w)} \) during the nonstandard expansion epoch, we obtain

\[ N \simeq 57 + \frac{1}{2} \ln \left( \frac{r}{0.1} \right) + \frac{1}{3(1+w)} \ln \left( \frac{\rho_{\text{RD}}}{\rho_{\text{end}}} \right) - \ln \left( \frac{\rho_{\text{RD}}^{1/4}}{10^{16} \text{GeV}} \right) - \ln \left( \frac{k}{0.05 \text{Mpc}^{-1}} \right) , \]

where \( \rho_{\text{RD}} \) is the total energy density at the time the standard RD epoch began, \( \rho_{\text{end}} = 3/(8\pi)^2 H^2(a_{\text{end}}) M_p^2 \) is the total energy density at the end of inflation, and we have fixed \( a_0 = 1 \), used \( T_0 = 2.725 \text{K} \), and assumed, for simplicity, that the transitions between different epochs were instantaneous\[6\]. Because quantities such as \( w, \rho_{\text{RD}} \) and \( \rho_{\text{end}} \) usually depend on the underlying particle physics model (see Sec. 2), the form of (44) is usually the most convenient choice to characterize the effect of post-inflationary evolution on models of inflation\[7\]. Furthermore, by assuming instantaneous thermalization among particles at the time the RD epoch began, one can connect the

\[6\] Note that our treatment does not account for non-instantaneous transitions between epochs reminiscent of e.g. nonperturbative effects encountered in preheating (see Sec. 2.1).

\[7\] See also Refs. [461–473] for scenarios where this reasoning was reversed so that the reheating temperature and the number of degrees of freedom were constrained by assuming a given model of inflation.
energy density at this epoch to the reheating temperature, \( \rho_{\text{RD}} = \pi^2 g_* T^4_{\text{ch}}/30 \). This is usually a safe assumption, as e.g. the Standard Model plasma generically thermalizes in much less than one e-fold from its production, see e.g. Ref. [179].

By assuming that the total energy density did not decrease much during the final e-folds of inflation\(^8\), one can show that the result (44) becomes

\[
N \simeq 57 + \frac{1 + 3w}{6(1 + w)} \ln \left( \frac{r}{0.1} \right) + \frac{1 - 3w}{3(1 + w)} \ln \left( \frac{\rho_{\text{RD}}^{1/4}}{10^{16} \text{ GeV}} \right) - \ln \left( \frac{k}{0.05 \text{ Mpc}^{-1}} \right),
\]

which is independent of \( \rho_{\text{end}} \) and only depends on \( r \) (which is an observable) and \( \rho_{\text{RD}} \). Note that the distance scale \( k_{\text{end}}^{-1} \) considered here is different from the current horizon \( 1/H_0 \) and therefore the number of e-folds (45) is not the one that solves the horizon and flatness problems. By choosing \( k = H_0 \simeq 0.0002 \text{ Mpc}^{-1} \) and assuming high scale inflation with instant reheating, we obtain \( N_{\text{horizon}} \simeq 62 \) in accord with Refs. [20, 21] (see also Ref. [22] for an upper limit on the number of e-folds, \( N \lesssim 68 \)).

Let us then give few examples of how post-inflationary expansion history affects model selection (for a larger review of different models, see Ref. [474]). In slow-roll approximation the number of e-folds between end of inflation and the horizon exit of a given scale can be connected to the scalar potential through

\[
N = \ln \left( \frac{\rho_{\text{end}}}{\rho_{\text{end}}} \right) = \int_{\phi_{\text{end}}}^{\phi_{\text{end}}} H(t)dt \simeq \frac{8\pi}{M_{\text{Pl}}^2} \int_{\phi_{\text{end}}}^{\phi_{\text{end}}} \frac{V}{V'}d\phi,
\]

where \( \phi_{\text{end}} \) is determined by \( \epsilon(\phi_{\text{end}}) = 1 \) and \( \phi_k \equiv \phi(ak) \). Therefore, for e.g. a power-law potential \( V \propto \phi^p \) one obtains at the leading order in the expansion in \( 1/N \) [21]

\[
n^p - 1 = -\frac{2 + p}{2N + p/2}, \quad r^p = \frac{16p}{p + 4N},
\]

where the superscripts refer to “power-law”. Another example is the (cosine) natural inflation model with a potential of the form \( V = \Lambda^4 (1 - \cos(\phi/f)) \), where \( \Lambda \) and \( f \) are mass scales determined by the underlying high energy theory [475] (see also Ref. [476] for other axion inflation models). The predictions of this model can also be calculated in the usual way from Eqs. (41), (42), and (46).

This type of potentials are examples of models where the predictions are relatively sensitive to the number of e-folds, and thus also to the details of post-inflationary expansion history. While the effect enters only through logarithmic corrections to \( N \) as in Eq. (44), our ignorance of the reheating energy scale \( \rho_{\text{RD}} \) allows their contribution to be substantial, and hence a nonstandard expansion phase can significantly affect model selection. However, there are also models where the value of \( r \) is relatively insensitive to \( N \). For example, the so-called \( \xi \) - and \( \alpha \)-attractor models [477-479] predict

\[
n^\alpha - 1 = -\frac{2}{N}, \quad r^\alpha = \frac{12\alpha}{N^2},
\]

where \( \alpha \ll N \) is a small parameter that is related to a pole in the kinetic term for the inflaton field. For example, for the Starobinsky model [2] \( \alpha = 1 \) and Linde-Goncharov model [480] \( \alpha = 1/9 \). Another notable example is Higgs inflation in metric gravity [138] which also corresponds to \( \alpha = 1 \), whereas for Higgs inflation in Palatini gravity \( \alpha \) can be much smaller than this [481], making \( r \) almost independent of \( N \) (although generally Palatini models are not \( \alpha \)-attractor models [482]). Also, in models where the gravity sector consists of a \( \beta R^2 \) term within Palatini gravity [483, 484], the prediction is \( r \simeq 1/\beta \) for large \( \beta \), i.e. practically independent of \( N \). As the energy scale of inflation is \( V^{1/4}/M_{\text{Pl}} \sim r^{1/4} \), this allows one to construct models where inflation occurs at a very low energy scale [485, 486].

In Fig. 7, we show the effect of a nonstandard expansion phase on inflationary observables for few example models for a broad range of e-folds achievable within nonstandard cosmology, see Eq. (45). We stress that the results only apply to scenarios where inflation was driven by a single scalar field that was in slow-roll at the time the measured perturbations were generated. The results can change substantially in the presence of e.g. a curvaton field (see Sec. 3.4) or in models where multiple fields took part in inflationary dynamics, see e.g. Refs. [488-490] for some recent work on the topic. Also, we note that the best-fit regime for \( n_s, r \) shown here was obtained from the Planck / Keck Array CMB data. Including other datasets can change the preferred values for these parameters, especially when the new dataset is in tension with the CMB measurements or when new physics is assumed to be involved. For example, the effects of nonstandard late-time expansion history (related to the so-called \( H_0 \) tension) have been discussed in Refs. [355, 491, 492] and the effects of nonstandard neutrino interactions and presence of extra radiation on inflationary model selection have been discussed in Refs. [493-496]. These aspects are discussed in more detail in Sec. 2.5 and Sec. 4.1.

\(^8\) Because the first slow-roll parameter is \( \epsilon = H/H^2 \ll 1 \), it is usually a very good approximation that the total energy density did not change much during the range of e-folds that is of interest here. In particular, this is the case for plateau models that give the best fit to the CMB data [13].
3.4. Curvaton scenario (Authors: C. Byrnes & T. Takahashi)

Remarkably, as discussed in the previous section, the statistical properties of the millions of temperature perturbation pixels observed in the CMB can be explained in terms of only two parameters – an amplitude and scale dependence – describing the primordial perturbations, Eq. (38). Extensive searches for additional primordial parameters, such as any form of non-Gaussianity, isocurvature perturbations, tensor perturbations or features in the primordial power spectrum have not detected any of these observables at high significance. Whilst these observations are consistent with some simple models of single-field inflation, it remains important to ask in which way the data could deviate from the simplest paradigm, and conversely, whether the absence of additional observables is evidence against any more complex model.

If only a single light field was present during inflation, then the perturbations today are necessarily adiabatic, meaning that on large scales, the Universe locally looks the same everywhere, with the energy density of all different components having a fixed ratio. In this case, each part of the Universe shares the same history and the (adiabatic) density perturbation can be seen as a time shift between different, super-horizon sized regions. In contrast, multifield inflation necessarily involves the existence of isocurvature perturbations during inflation because the fields will fluctuate differently and hence local patches will contain different ratios of each scalar field. Whether these isocurvature perturbations persist until the present day is model dependent, with knowledge (or assumptions) of the entire reheating history required, from the end of inflation until the primordial components of baryons, DM, neutrinos and radiation have been produced.

3.4.1. Curvaton perturbations and local non-Gaussianity

The curvaton scenario [497–499] (see also Refs. [500, 501] for related older works) is arguably the most popular model that can generate large non-Gaussianity [381, 502–520] and/or isocurvature perturbations [521–524], and it relies on the existence of a late decaying scalar field (e.g. a modulus field [499]) and is often connected to an EMDE. Both of these observational signatures are potential smoking gun signatures of multiple field inflation, and hence the curvaton scenario is a popular test case for studying inflation with more than one light scalar field present during inflation. The existence of multiple light fields is arguably natural in the context of high energy theories such as supersymmetric theories and string theory (see e.g. Ref. [525]). Even if one does not accept this motivation, unless the SM Higgs was the inflaton field (which requires a very large nonminimal coupling to gravity), at least two scalar fields were present during inflation.

The curvaton field $\sigma$ is posited to be a light and energetically subdominant field during inflation, which typically has essentially no impact on the background dynamics during inflation. However, provided that the curvaton mass is light compared to the Hubble parameter during inflation ($m_{\sigma} \ll H$), the curvaton field will be perturbed and its perturbations will form a quasi-scale invariant spectrum. In the simplest curvaton model, the curvature perturbation is given by

$$\zeta = \frac{2}{3} r_{\text{decay}} \frac{\delta \sigma_{x}}{\sigma_{x}},$$

where $r_{\text{decay}} \equiv 3\rho_{\sigma}/(4\rho_{\nu} + 3\rho_{\sigma})|_{\text{decay}}$ models the fraction of the background energy density in the curvaton field at the
time it decays, with $\rho_\sigma$ and $\rho_r$ the energy densities of the curvaton and radiation evaluated at the time of the curvaton decay, respectively, and $\sigma_\epsilon$ and $\delta_\epsilon$ are the curvaton field and its fluctuations at the time of horizon exit of the scale of interest. In order for the curvature perturbations generated from the curvaton to become a non-negligible part of the primordial curvature perturbation, for example observed via the CMB temperature perturbations, the background energy density of the curvaton must become a significant – or even dominant – fraction of the background before it decays, since $r_{\text{dec}} \gg 10^{-5}$ given $\delta\sigma_\epsilon/\sigma_\epsilon \ll 1$.

Provided that the inflaton field decays first into a radiation bath, the background energy density of the Universe decays like $a^{-4}$ while the curvaton’s energy density is initially frozen until $H \sim m_\sigma$, and afterwards decreases like $a^{-3}$ while it oscillates in a quadratic minimum [498, 499], until the curvaton also decays into radiation. It is therefore a standard requirement of the curvaton scenario that it decays much more slowly than the inflaton field and that its potential has a quadratic minimum.

The curvaton model is a popular way of generating local non-Gaussianity, which measures the correlation between long and short-wavelength modes. It is defined via the bispectrum (related to the 3-point correlation function) $B_\zeta$ as

$$\langle \zeta(k_1)\zeta(k_2)\zeta(k_3) \rangle = (2\pi)^3\delta^3(k_1 + k_2 + k_3) B_\zeta(k_1, k_2, k_3),$$

where the local bispectrum shape is related to local $f_{\text{NL}}$ by

$$B_\zeta^\text{local} = \frac{6}{5} f_{\text{NL}} (P_\zeta(k_1)P_\zeta(k_2) + 2 \text{ permutations}).$$

In the simplest case, the curvaton is assumed to have a quadratic potential and negligible couplings to the inflaton field (any couplings could allow the inflaton to decay into the curvaton, which can have a big phenomenological impact given the initial subdominance of the curvaton [506, 526–528]). In this case, analytic results exist for the spectral index of the curvaton perturbations and the local non-Gaussianity in terms of $r_{\text{decay}}$:

$$n_s - 1 = -2\epsilon + 2\frac{m_a^2}{3H_*^2},$$

$$f_{\text{NL}} = \frac{5}{4r_{\text{decay}}} - \frac{5}{3} - \frac{5r_{\text{decay}}}{6},$$

where $\epsilon = -\dot{H}/H^2$ is a slow-roll parameter which is determined by the inflaton potential. In the expression for $n_s$, $H_*$ is the Hubble parameter at the time of horizon exit. Current constraints on the spectral index, which require $n_s < 1$ at high significance, require the inflaton field to be of the large-field variety with sufficiently large $\epsilon$ since $m_a^2/H_*^2$ always gives a positive contribution to the spectral index. This is the case for the simplest curvaton case. However, if one considers an axion-type potential for the curvaton or a self-interaction, the above prediction for $n_s$ and $f_{\text{NL}}$ are modified [510–515, 518–520].

Therefore, despite the inflation field perturbations being subdominant (and perhaps completely unobservable, but see Refs. [199, 200, 529–536] for mixed models in which both the inflaton and curvaton perturbations are important), the inflaton potential remains quite tightly constrained, and for example, the “simplest” curvaton scenario where both the inflaton and curvaton fields have quadratic potentials is now observationally excluded [205, 537]. However, a quartic inflaton plus quadratic curvaton provides an excellent fit to the data unlike in the single-field case [533], since, when the curvaton is a dominant source of density fluctuations, the tensor-to-scalar ratio is very suppressed [199, 200, 529, 531] even if one assumes an inflaton potential that would predict a too large tensor-to-scalar ratio in the case that it also generated the observed perturbations. Alternatively, the spectral index can be made red for small-field models of inflation if the curvaton potential is more complex and its effective mass is negative during inflation [538].

Again considering the quadratic curvaton, one can see from Eq. (53) that large non-Gaussianity is possible if $r_{\text{decay}} \ll 1$. However, the current Planck non-Gaussianity constraint of $|f_{\text{NL}}| \lesssim 10$ already requires $r_{\text{decay}} \gtrsim 0.2$ [14] assuming that the curvaton perturbations dominate, so the curvaton must have at least come close to dominating the background energy density of the Universe by the time it decays. Given the short additional time it takes for the curvaton to dominate, $f_{\text{NL}} \approx -5/4$ is a rather natural prediction of the curvaton scenario, independently of the inflaton potential [208, 539]. A detection of this value of $f_{\text{NL}}$, which may be marginally within reach of sky galaxy surveys such as SPHEREX, Euclid, SKA and so on within a decade [540–545], would be strong evidence that the Universe underwent an EMDE caused by curvaton domination.

By measuring additional non-Gaussianity parameters it is possible to test two of the key assumptions underlying the simplest curvaton scenario, namely whether the curvaton has a quadratic potential and whether all the perturbations were purely generated by a single curvaton field (see Refs. [546, 547] for introductory references). The latter case can be tested by measuring the two trispectrum (4-point correlation function) parameters, $\tau_{\text{NL}}$ and $g_{\text{NL}}$, with $g_{\text{NL}}$ defined by

$$\zeta(x) = \zeta_G(x) + \frac{3}{5} f_{\text{NL}} \left( \zeta_G^2(x) - \langle \zeta_G^2(x) \rangle \right) + g_{\text{NL}} \zeta_G^3,$$

where $\zeta_G$ is the local Gaussian fluctuations and $\langle \cdot \rangle$ is the spatial average within comoving Mpc-scale superhorizon volume.
while $\tau_{NL}$, which has a different shape dependence than $g_{NL}$, is defined in e.g. Refs. [548–550]. In the case that the perturbations of a single field generate the curvature perturbation, $\tau_{NL}$ must satisfy the consistency relation $\tau_{NL} = (6 f_{NL}/5)^2$, whilst in general it will be larger [551, 552]. The quadratic curvaton predicts $g_{NL} \sim f_{NL}$ which is unobservably small, but nonquadratic models can (with tuning) produce $|g_{NL}| \gg |f_{NL}|$, as well as a potentially large scale-dependence of $f_{NL}$ [362, 532, 553–556]. However, given the observational difficulty in determining the trispectrum parameters and/or the scale dependence of $f_{NL}$ and given the current constraint on $f_{NL}$, these higher-order non-Gaussian parameters are only expected to become detectable in quite a limited range of parameter space where the curvaton potential deviates significantly from being quadratic and/or in cases where it generates highly non-Gaussian perturbations but the inflaton perturbations dominate at the linear level [557].

Before closing this section, we make some comments on the relation of the curvaton to the SM Higgs field. Although one may be tempted to consider the curvaton field as the Higgs, the Higgs curvaton would not work within the standard thermal history case [558–560] since the Higgs has a quartic potential, implying the oscillating curvaton energy density scales as $\rho \propto a^{-4}$, making it impossible to have a large enough fraction of the energy density in the curvaton field to generate significant curvature perturbations. Allowing an early period of kination alleviates this problem but it remains extremely challenging to make the Higgs work as a curvaton and/or provide the dominant energy source of reheating, even when allowing a nonminimal coupling to gravity [472, 558, 560, 561]. The case where the curvaton is coupled to the Higgs has also been studied and in that case it may leave some observable signature [562, 563].

### 3.4.2. Isocurvature perturbations and the evolution of non-Gaussianity after inflation

The survival of isocurvature perturbations is not generic, because a complete thermalisation of the Universe during reheating, with no conserved quantum numbers, completely converts the initial isocurvature perturbations into adiabatic perturbations [564]. Given the great uncertainty about how reheating proceeded, the time at which DM was created, the time of baryogenesis, and so on, no strong statements can be made about whether the persistence of isocurvature perturbations should be considered likely or fine-tuned. Hence, the observational absence of any variety of isocurvature perturbations cannot be – in general – interpreted as evidence against the existence of multiple light scalar fields during inflation [565, 566].

If isocurvature perturbations were observed they could potentially provide many additional observables to probe both inflation and reheating, albeit with some degeneracy between the two. The Planck collaboration have searched for many types of isocurvature perturbations, including baryonic, DM, neutrino density and neutrino velocity isocurvature perturbations, in all cases measured relative to the primordial radiation curvature perturbation, and obtained severe constraints on all of these types [13]. In all cases, these isocurvature perturbations can be partially or fully correlated (or anti-correlated) with the adiabatic perturbations, and they may also have a very different scale dependence. If all the fields are slowly-rolling during inflation it is reasonable to expect the isocurvature perturbations to also be quasi-scale invariant. However, if the fields that did not take part in driving inflation are in a regime where stochastic quantum fluctuations, rather than the classical slow-roll, dominated their evolution during inflation, the resulting isocurvature spectrum can be strongly scale-dependent [383, 567–569].

In the curvaton scenario, isocurvature fluctuations can also be generated. For example, when DM freeze-out (or production of DM) and/or baryogenesis occur before the curvaton decay, large (anti-correlated) baryon/DM isocurvature fluctuations can be produced [381, 509, 521–524], which are excluded by current data unless the curvature perturbation is also generated from the inflaton sector [521, 523, 524]. We note that the issue of isocurvature fluctuations in the curvaton scenario should be carefully investigated especially when the curvaton decays after it dominates the Universe [570], which may well be the case given the observational constraint on $f_{NL}$ [14]. When DM and/or baryons are generated from the curvaton decay, correlated isocurvature fluctuations whose size depends on the $r_{dec}$ parameter can be generated [381].

Non-Gaussianity of isocurvature fluctuations have also been studied [571–574], which can also be useful to constrain models with isocurvature fluctuations, including the curvaton scenario. In general, constraints from power spectrum observations are tighter than those from non-Gaussianities, however, in some particular cases non-Gaussianities can give a competitive or more severe constraint (for instance, see Refs. [575, 576] for the case of axion DM and the curvaton). Since isocurvature fluctuations can be correlated with adiabatic ones, several types of $f_{NL}^{(iso)}$ are defined and constraints on those have been investigated by using WMAP data [575–577] and Planck data [14]. Their expected constraints from future observations have been studied too [578, 579]. For discussion on the trispectrum from isocurvature fluctuations, see Refs. [580, 581].

One of the most important impacts of isocurvature perturbations, whether or not they persist until the present day, is that their existence allows the curvature perturbation to evolve on super-horizon scales [582–584]. Therefore, even the predictions of the standard adiabatic perturbations may evolve and therefore, in contrast to single-field inflation, one cannot just match the predictions of the curvature perturbation calculated at horizon exit during inflation to the primordial curvature perturbation. This entails a significant additional complexity, rendering all predictions potentially dependent on the details of the full inflationary history and reheating. The quadratic curvaton scenario represents a simple case study, a model where initially the curvature perturbation is Gaussian and generated by the inflaton, but as the late-decaying curvaton field starts to dominate the background energy density its perturbations may also dominate
and the perturbations become non-Gaussian (and conserved after the curvaton field has decayed). In general, making analytic predictions is hard or even impossible, but extensive work has been done in the case of perturbative reheating following multiple (normally two-field) inflation at the spectral index, bispectrum and trispectrum [74, 585–590]. Some reasonably model-independent conclusions are that if there is non-negligible non-Gaussianity present at the end of inflation, then its amplitude often changes by an order unity factor until the end of reheating (although potentially switching sign), while the spectral index is limited to lie in between the values set by the smallest and largest values of the spectral indices of the individual scalar fields.

3.5. Formation of microhalos (Author: A. Erickcek)

As discussed in previous sections, the expansion history of the Universe affects the evolution of curvature perturbations and the eventual growth of structure. For adiabatic perturbations, deviations from the standard post-inflationary epoch of radiation domination will only affect perturbations on scales that lie within the cosmological horizon during the period of nonstandard evolution. The requirement that the Universe be radiation dominated at the time of neutrino decoupling [23–27, 259] ensures that these scales are smaller than 200 pc [43]. Since the signatures are confined to such small scales, only dark matter perturbations possibly retain a record of deviations from radiation domination in the early Universe. Therefore, in this section, we will be primarily interested in the evolution of perturbations in the density of dark matter and how that evolution leaves an imprint on the matter power spectrum on sub-kiloparsec scales. Enhancements to the small-scale dark-matter power spectrum will lead to earlier-forming and more abundant gravitationally bound structures. The amount of dark matter present within the cosmological horizon at the time of neutrino decoupling implies that the masses of these structures are less than 1200 $M_{\odot}$, and an earlier onset of radiation domination would confine the enhanced structures to even smaller masses [43]. Earth-mass halos are frequently called microhalos; we adopt that terminology here and extend it to include even smaller halos.

3.5.1. Growth of density perturbations

The evolution of dark matter density perturbations is determined by four factors: the expansion rate of the Universe, the evolution of curvature perturbations, interactions between the dark matter and other species, and the random motions of dark matter particles. The first two factors are determined by the properties of the Universe’s dominant component and can readily enhance the growth of dark matter perturbations, while the latter two factors depend on the properties of the dark matter particle and generally suppress the growth of dark matter density perturbations.

The expansion rate of the Universe and the evolution of curvature perturbations are both determined by the pressure of the dominant component of the Universe. If the Universe’s dominant component is pressureless, then its density scales as $\rho \propto a^{-3}$, and the metric perturbation in Newtonian gauge $\Phi$ remains constant as perturbation modes enter the cosmological horizon [17]. For perturbation modes with comoving wavenumber $k$, Poisson’s equation states that $(k^2/a^2)\Phi = 4\pi G \rho \delta$, where $\delta$ is the density fluctuation. Since $\rho \propto a^{-3}$, the fact that $\Phi$ is constant implies that $\delta$ grows linearly with the scale factor. The same perturbation evolution applies to oscillating scalar fields with quadratic potentials. Since the pressure averages to zero over many oscillations, perturbations with $k/a < \sqrt{3H m_{\phi}}$, where $m_{\phi}$ is the mass of the scalar field $\phi$, grow linearly with the scale factor [129, 130] and can even form bound structures [80, 591, 592]. Furthermore, scalar fields with potentials that become shallower than quadratic away from their minima quickly form localized structures called oscillons that behave like massive particles [79, 119]. Therefore an EMDE characterized by $\rho \propto a^{-3}$ and constant metric perturbations on subhorizon scales may be caused by massive particles or scalar fields. In both cases, transitioning to radiation domination requires the dominant component of the Universe to decay into relativistic particles, at which point the subhorizon metric perturbations undergo decaying oscillations [43].

If dark matter does not interact with other particles, then the Boltzmann equation that governs the evolution of dark matter perturbations only depends on expansion rate and the metric perturbation $\Phi$, see e.g. Ref. [47]:

$$a^2\delta_{DM}' + \left[3 + \frac{aH'}{H}\right]a\delta_{DM}' = \frac{k^2}{a^2 H^2} \Phi - 3 \left[3 + \frac{aH'}{H}\right] a\Phi' - 3a^2\Phi'' ,$$

where the prime denotes differentiation with respect to the scale factor $a$. During an EMDE, $\Phi$ is constant and $H(a) \propto a^{-3/2}$, in which case Eq. (55) implies that $\delta_{DM}$ is constant on superhorizon scales ($k \ll aH$) and grows linearly with the scale factor on subhorizon scales ($k \gg aH$). This evolution is not altered if the dark matter is produced by the decay of the dominant component during the EMDE [43]. If dark matter is initially in chemical equilibrium with relativistic particles and then freezes out, dark matter annihilations alter the growth of perturbation modes that enter the horizon prior to freeze-out, but the dark matter particles’ subsequent fall into the gravitational wells created by growing perturbations in the dominant component washes out this effect. If the freeze-out temperature is greater than twice the reheating temperature, there is no significant change in $\delta_{DM}$ at the end of the EMDE compared to the noninteracting case [340]. If dark matter is produced by the decay of the dominant component and then self-annihilates at the end of the EMDE, $\delta_{DM}$ decreases by roughly a factor of ten at reheating relative to the noninteracting case due to the higher annihilation rate in overdense regions [45]. A similar suppression occurs if dark matter is co-decaying
For modes that enter the horizon well before the end of the EMDE \((k \gtrsim \sqrt{\Omega_D a_{reh} H_{reh}})\), where \(a_{reh}\) and \(H_{reh}\) are the scale factor and Hubble rate, respectively, when the Universe becomes RD, the reduction in \(\delta_{DM}\) due to annihilations at reheating is less than the growth of the \(\delta_{DM}\) during the EMDE, implying that an EMDE still enhances the amplitude of density perturbations on these scales.

If the dominant component of the Universe has significant pressure, \(\Phi\) rapidly decays after the perturbation mode enters the horizon [47]. The right-hand-side of Eq. (55) becomes negligible, implying that

\[
\delta_{DM} \propto \int^a \frac{da}{a^3 H(a)}.
\]

This solution has a simple physical interpretation. Immediately after a perturbation mode enters the horizon, dark matter particles are gravitationally pulled toward overdense regions. When \(\Phi\) decays, the particles stop accelerating and drift toward the initially overdense region with physical velocity \(v \propto 1/a\). The comoving displacement of these particles is \(\int v dt/a = \int v da/(a^2 H)\), and the convergence and divergence of these displacements results in a growing dark matter density perturbation. Equation (56) recovers the well-known result that \(\delta_{DM}\) grows logarithmically with scale factor while the Universe is radiation dominated, but it also applies to any expansion phase that includes vanishing gravitational perturbations. If the Universe is dominated by an energy source that is stiffer than radiation, then \(\delta_{DM}\) will grow faster than it does during radiation domination: \(\delta_{DM} \propto a^{(3w-1)/2}\), where \(w\) is the equation of state parameter for the dominant component of the Universe. Therefore, any expansion phase with \(w > 1/3\) will enhance dark matter density perturbations on subhorizon scales. An extreme example is a period of kination, during which the Universe is dominated by a fast-rolling scalar field \((w \approx 1)\) and \(\delta_{DM}\) grows linearly with the scale factor [47]. Conversely, an expansion phase with \(0 < w < 1/3\) will slightly suppress subhorizon perturbations because these modes will stagnate instead of growing logarithmically with the scale factor after horizon entry.

The impact of a nonstandard expansion phase on the matter power spectrum depends on the evolution of both \(\delta_{DM}(a)\) and \(H(a)\). The linear growth of \(\delta_{DM}(a)\) during both an EMDE and a period of kination implies that \(\delta_{DM}\) at the onset of radiation domination will be proportional to \(a_{reh}/a_{hor}\), where \(a_{hor} = k/(H_{hor})\) is the value of the scale factor when the mode enters the horizon. During an EMDE, \(a_{hor} \propto k^{-2}\), whereas \(a_{hor} \propto k^{-1/2}\) during kination. In both cases, \(\delta_{DM}(k)\) is also proportional to the value of \(\Phi(k)\) while the mode is outside the horizon, which has a dimensionless power spectrum proportional to \(k^{n_{s}-1}\). It follows that the dimensionless power matter power spectrum \(\Delta^2(k)\) will be proportional to \(k^{3+n_{s}}\) for modes that enter the horizon during an EMDE and will be proportional to \(k^{n_{s}}\) for modes that enter the horizon during kination. For a general expansion phase with \(w > 1/3\), \(\Delta^2(k) \propto k^{n_{s}+(3w-3)/(3w+1)}\) for modes that enter the horizon during this period. All of these power spectra imply enhanced inhomogeneity on small scales compared to the \(\Delta^2(k) \propto k^{n_{s}-1} \ln^2[k/(8k_{eq})]\) scaling for modes that enter the horizon during radiation domination [594].

### 3.5.2. Cutting off microhalo formation

Since \(\Delta^2(k)\) increases as \(k\) increases for modes that enter the horizon during an EMDE or during an expansion phase with \(w > 1/3\), smaller scales are increasingly inhomogeneous and could collapse to form bound structures at arbitrarily early times. It is not possible for \(\Delta^2(k)\) to continue monotonically increasing with \(k\) indefinitely, however. As an extreme limit, the power spectra derived above do not apply to modes that are so small that they never exit the horizon during inflation. In most cases, the growth of the matter power spectrum cuts off at much larger scales than the horizon size at the end of inflation due to the dark matter particles’ thermal motions or their interactions with other particles.

If the dark matter interacts with relativistic particles, then perturbations that enter the horizon while the dark matter is kinetically coupled to these relativistic particles will experience damped oscillations between horizon entry and the decoupling of the dark matter, implying that \(\Delta^2(k)\) decreases with increasing \(k\) for such modes, see e.g. Ref. [595]. There is an exception to this generic behavior during an EMDE, however. During an EMDE, the dominant matter component decays into relativistic particles, and once any pre-EMDE radiation has sufficiently diluted, the energy density in these particles scales as \(\rho_r \propto a^{-3/2}\) due to the production of new relativistic particles [244]. At this point, density perturbations in these relativistic particles do not oscillate, but rather grow until they reach a steady state during which the production of new relativistic particles in increasingly overdense regions is balanced by the outflow of relativistic particles from these regions [43, 45]. If dark matter is in kinetic equilibrium with the relativistic particles, the dark matter particles are pulled out of the overdense regions by the outflow of relativistic particles, and the dark matter accumulates in underdense regions. If the dark matter remains in kinetic equilibrium through reheating, the dark matter remains concentrated in these regions, resulting in large isocurvature perturbations on scales that enter the horizon during the EMDE [596]. If the dark matter kinetically decouples prior to reheating, the dark matter particles quickly fall into the gravitational wells created by density perturbations in the dominant component, and the dark matter power spectrum is not significantly different from the noninteracting case [597]. Therefore, interactions between dark matter particles and the relativistic particles produced during an EMDE do not suppress perturbations on scales that enter the horizon prior to kinetic decoupling, and perturbations on these scales are still significantly enhanced by an EMDE.
After dark matter particles kinetically decouple from all other particles, their random thermal motions allow them to free stream out of overdense regions. Perturbations on scales smaller than the comoving free-streaming horizon are exponentially suppressed, with $\Delta^2(k) \propto e^{-k^2/k_{fs}^2}$, where $k_{fs}$ is the inverse of the comoving free-streaming horizon, see e.g. Ref. [505]. If dark matter is a decay product of the species that dominates the Universe during an EMDE and never interacts with other particles, its free streaming will inhibit the formation of microhalos unless its initial velocity is much less than the speed of light [43, 45, 598]. If dark matter interacts with other particles, the free-streaming horizon is determined by the velocity of the dark matter particles when they kinetically decouple. During a nonstandard expansion phase, the Hubble rate is larger than it is during a radiation-dominated phase with the same temperature, so DM particles will decouple from SM particles at a higher temperature than they would during radiation domination [599–601]. Since the velocities of dark matter particles decrease as $v \propto 1/a$ after decoupling, an earlier decoupling reduces the free-streaming horizon. If dark matter kinetically decouples during an EMDE, the free-streaming horizon is further reduced because the continual production of relativistic particles during an EMDE slows their cooling, leading to a greater reduction in the dark matter particles’ velocities over a given temperature range. Gelmini and Gondolo [599] showed that the minimum halo mass set by the free-streaming horizon can be reduced by several orders of magnitude if the dark matter kinetically decouples during an EMDE.

Kinetic decoupling during an EMDE is more complicated than Gelmini and Gondolo [599] realized, however. If dark matter is in kinetic equilibrium with the relativistic particles that are created during an EMDE, the slow cooling of these particles implies that the dark matter is still heated by its interactions with these particles long after the momentum transfer rate falls below the Hubble rate, which is the usual criterion for kinetic decoupling [602]. The standard method of calculating the dark matter temperature [595, 603] indicates that this “quasi-decoupled” state persists until reheating, but this approach assumes that each collision causes a small fractional change in the dark matter particles’ momenta. During an EMDE, this assumption ceases to hold at about the same time that the Hubble rate equals the collision rate [597], which is roughly $m_{DM}/T$ times the momentum transfer rate [603]. Since we do not know how the temperature of the dark matter evolves after this time during an EMDE, we cannot calculate the free-streaming horizon if dark matter kinetically decouples from SM particles during an EMDE.

If the nonstandard expansion phase lasts long enough for $\delta_{DM}$ to grow larger than one, it is possible to form microhalos prior to reheating. Structure formation during an EMDE would progress similarly to structure formation during matter domination: gravitational forces cause overdense regions to break away from the Hubble flow and collapse into bound structures when the value of $\delta$ predicted by linear theory exceeds 1.686. If the collapsing region is sufficiently spherical and homogeneous, it will form a PBH, as will be discussed in Sec. 3.6. The more common outcome of gravitational collapse is the formation of a virialized microhalo that is primarily composed of whatever matter-like substance dominates the Universe during the EMDE, with just a small portion of the mass coming from dark-matter particles. When the dominant component during the EMDE decays, these halos will become unbound, and the dark matter particles within these halos will be released. These particles have higher velocities than unbound particles, and even more importantly, the directions of these velocities are randomized. The subsequent free streaming of these gravitationally heated particles erases the perturbations that grew significantly during the EMDE, thus preventing the late formation of microhalos [341]. Structure formation while the Universe is dominated by a component with $w \geq 1/3$ is less understood because the dark matter particles are merely drifting towards each other, but Blanco et al. [341] demonstrated that it is possible to form dark matter halos during radiation domination if the convergence of the drifting particles causes a region to become locally matter dominated. Unlike the halos that form during an EMDE, these halos would be composed solely of dark matter particles and would survive the transition to radiation domination.

### 3.5.3. The microhalo population

The first microhalos will form on scales with the largest values of $\Delta^2(k)$, therefore the peak scale in the power spectrum sets the mass of the first microhalos: $M \simeq (4\pi/3)\rho_{DM,0}k_{peak}^{-3}$, where $\rho_{DM,0}$ is the present-day DM density. Roughly speaking, these halos will form when $\Delta^2(k_{peak}) > 1$. For halos that form during the matter-dominated era, when $\Delta^2(k) \propto a^2$, the typical value of the scale factor when halos form is proportional to $1/\sqrt{\Delta^2(k)}$. Since $\Delta^2(k)$ increases very slowly with $a$ for modes that enter the horizon during radiation domination, a nonstandard expansion phase reduces the typical value of the scale factor at halo formation by roughly a factor of $\sqrt{\Delta^2(k_{reh})}/\Delta^2(k_{peak})$, where $k_{reh} = a_{reh}H_{reh}$. The peak scale is determined by the cut-off scale in the matter power spectrum, so the formation time of the first halos is most sensitive to the ratio $k_{cut}/k_{reh}$, with only a weak dependence on the reheat temperature [343]. The formation time of a halo determines its density, and Delos et al. [604] demonstrated that the density profiles of these first halos can be predicted directly from $\Delta^2(k)$.

Looking beyond the first halos, the halo abundance predicted by Press-Schechter theory [605] is enhanced for all scales with enhanced $\Delta^2(k)$. The number density of such halos depends on their masses, but the total fraction of dark matter contained in halos at a given time is largely independent of $k_{reh}$ and instead depends on the maximal enhancement of $\Delta^2(k)$, as determined by $k_{cut}/k_{reh}$, see e.g. Ref. [340]. Following a nonstandard expansion phase that

---

9 If dark matter is co-decaying [306], then the halos will be nearly half dark matter, but the vast majority of these particles will be annihilated during reheating.
3.6. Formation of primordial black holes (Authors: T. Harada & K. Kohri)

Primordial black holes (PBHs) may have formed in the early Universe [606, 607]. In the present Universe, PBHs can act as gravitational sources such as DM, or as GW sources. Furthermore, since the mass of the PBHs can be very small and they can lose their mass by the Hawking radiation, PBHs provide us with a unique probe into quantum gravity effects of black holes. For this reason, PBHs have attracted intense attention not only in cosmology but also in other areas of fundamental physics.

In previous sections, we have discussed mechanisms to produce curvature or density perturbations by models of inflation and curvaton. Because the quantum generation of fluctuations in such scenarios implies a small but nonvanishing probability of large amplitude of perturbations, the existence of PBHs is quite realistic. Therefore, an interesting question is how many PBHs could have formed in the early Universe. In this section, we discuss PBHs formed through the collapse of overdense regions due to large density perturbations at small scales. For a recent review paper including other mechanisms for PBH formation, we refer to Ref. [93].

3.6.1. Radiation-dominated universe

In the early Universe, naively one would expect that the RD Universe was realized just after inflation, where the radiation temperature was larger than $O(1)$ MeV to be consistent with BBN (see Sec. 4.1). To theoretically predict the abundance of PBHs, it is important to identify the conditions for PBH formation in terms of the amplitude of perturbations. A region with a large density perturbation $\delta$ which is enclosed by a length scale $R$ can collapse into a PBH. In 1975, B. J. Carr proposed a criterion by which the overdense region becomes a black hole if $R$ is larger than the Jeans radius at the time of its maximum expansion [608]. His criterion gives $\delta > \delta_c \sim c_s^2$ with the sound speed $c_s^2 = w$ where $w$ is the EoS parameter. In a RD Universe, the criterion is $\delta_c \sim 1/3$. With more sophisticated methods in general relativity, threshold values in the range $\delta_c \approx 0.4 - 0.7$ have been reported. For example, spherically symmetric numerical relativity simulations report $\delta_c \approx 0.42 - 0.66$ [609–612]. Analytically, the authors of Ref. [613] derived the formula

$$\delta_c = \frac{3(1 + w)}{5 + 3w} \sin^2 \left(\frac{\pi\sqrt{w}}{1 + 3w}\right),$$

which gives $\delta_c \approx 0.41$ in a RD Universe. However, it is worth remarking that the threshold actually depends on the spatial profile of the overdense region [609–618]. It should also be noted that near the threshold, there appear critical phenomena, such as the scaling law of the black hole masses and the self-similar solution at the threshold [610, 616, 619], which were originally discovered in the Einstein-scalar field system in an asymptotically flat spacetime [620]. The scaling law of PBH masses significantly deforms the observational bound on the power spectrum of the perturbation because the PBH mass can be much smaller than the mass enclosed within the Hubble horizon at the horizon entry of the perturbation [621, 622].

To obtain the abundance of PBHs as a function of density perturbation or a variance $\sigma^2$ of it, for simplicity we can follow the Press-Schechter formalism [605]. A PBH can form just after a region with $\delta > \delta_c$ enters the Hubble horizon. By assuming that $\delta$ obey a Gaussian distribution with the variance $\sigma^2$, the probability of the PBH formation is given by

$$\beta = \int_{\delta_c}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{\delta^2}{2\sigma^2} \right) d\delta \simeq \text{erfc} \left( \frac{\delta_c}{\sqrt{2\sigma}} \right) \approx \sqrt{\frac{2}{\pi}} \frac{\sigma}{\delta_c} \exp \left( -\frac{\delta_c^2}{2\sigma^2} \right).$$

It is notable that $\beta$ is nothing but $\rho_{\text{PBH}}/\rho_{\text{tot}}$ at the time of formation of the PBHs where $\rho_{\text{PBH}}$ and $\rho_{\text{tot}}$ are the energy densities of PBHs and the background, respectively. In Fig. 8, we plot this RD case with a green curve.

For simplicity, we show an approximate relation between the mass $m_{\text{PBH}}$ and the cosmic time at the formation $t_{\text{form}}$ for a PBH formed during a RD epoch:

$$m_{\text{PBH}} \sim m_p^2 t_{\text{form}} = \frac{m_p^2}{T_{\text{form}}^2} \sim 10^{15} g \left( \frac{T_{\text{form}}}{3 \times 10^9 \text{GeV}} \right)^{-2} \sim 30 M_\odot \left( \frac{T_{\text{form}}}{40 \text{MeV}} \right)^{-2},$$

where $m_p = M_p/\sqrt{8\pi} \simeq 2.4 \times 10^{18} \text{GeV}$ is the reduced Planck mass and $T_{\text{form}}$ is the radiation temperature at the time of the PBH formation.

Compared to the above method, more sophisticated ones have been reported by Refs. [624, 625] for peak statistics, Ref. [626–628] for choices of the window function, Refs. [615, 624, 629] for nonlinear relations between curvature and density perturbations and Refs. [630–633] for possible extensions of peak theory.

significantly enhances $\Delta^2(k)$ on small scales, over half of the dark matter can bound into microhalos at redshifts well above 100, long before structure is expected to form in standard cosmologies [43, 340]. It is even possible to form microhalos prior to matter-radiation equality [341].
Next we discuss formation of PBHs during an EMDE. As was discussed in the previous sections, in modern cosmology it is known that an EMDE may have been realized just before the onset of the standard RD Universe. Considering the formation of PBHs during an EMDE, a severe problem is that there exist two conflicting effects: i) The EoS parameter $w$ is nearly zero since the cosmic pressure is negligible. Thus, as we can see in Eq. (58) which gives $δ = 0$, more PBHs could be produced. ii) On the other hand, density perturbation can evolve during a MD epoch, which means that also anisotropies of the density contrast should have evolved in the 3D space. In this case, a collapsed region may not be easily enclosed by the (event) horizon.

To calculate the formation rate of PBHs during a MD epoch, in Ref. [53] we have applied the Zel’dovich approximations to the collapsing objects anisotropically in the 3D space. To be a PBH, we judged if the region is enclosed by the apparent horizon by adopting the Hoop Conjecture for such non-spherical object. Due to a finite angular momentum induced by the anisotropic collapse, the abundance of PBHs produced in this case is exponentially suppressed [55]. This means that a produced PBH tends to be spinning nearly extremely. In other words, a more spinning object than an extremal one cannot become a PBH, but instead a kind of halo such as a CDM halo around a galaxy should form, as discussed in Sec. 3.5.

In Fig. 9, we plot the spin distribution of PBHs formed in a MD epoch due to the first-order effect, discussed in Ref. [55]. The $x$-axis is the normalized Kerr parameter $a_* = L/(Gm_{\text{PBH}}^2/c^2)$ where $L$ is the angular momentum of the PBH. The curves denote the spin distribution functions normalized by their maximum values for the variance of density fluctuations, depending on $σ$. In Fig. 8, we instead plot the abundance of PBHs formed during a MD epoch as

**Fig. 8.** The PBH abundance $β$ as a function of the mean variance $σ$ of the density perturbation $δ$. The cases of the radiation domination (here: RD) [613] and the early matter domination (here: MD) for both the first- and second-order effects [55] are plotted by the green and two red curves, respectively. The line of the power law $= 0.0556\sigma^5$ denotes the one without suppression by finite angular momentum in case of the early MD studied by Refs.[50, 53, 623]. The curves are from Ref. [55].

**Fig. 9.** Spin distribution of PBHs formed during an EMDE due to the first-order effect (left) and the second-order effect (right) [55]. The $x$-axis is the normalized Kerr parameter $a_* = L/(Gm_{\text{PBH}}^2/c^2)$ with $L$ being the angular momentum. The curves denote the spin distribution functions normalized by their maximum values for the variance of density fluctuations, depending on $σ = 0.1, 0.05,$ and 0.01. This figure is from Ref. [55].
We start by discussing the BBN constraints. BBN represents the earliest epoch of the Universe from which we have data. The agreement between the observed primordial abundances of helium and deuterium with the predicted ones is a highly nontrivial success of the standard cosmological model. Equivalently, BBN serves as a powerful probe of new physics and nonstandard cosmological expansion histories.

In this section we will discuss how observations of the CMB and the primordial element abundances can be used to obtain information about both the early and late (pre- and post-BBN) expansion history of the Universe, and how the topic has been discussed in the literature. To illustrate the constraining power of current cosmological observations, we outline in Table 2 the up to date CMB and BBN constraints on key parameters constraining nonstandard expansion histories.

We also refer to Ref. [56] for another effect, which allows to simply multiply the expression for $\beta$ by an additional power $\propto \sigma^{3/2}$ induced by the initial inhomogeneity of the Universe. However, a complete analysis of generation of finite spin and suppression of $\beta$ with an initial inhomogeneity has not been fully done. Therefore, for the moment, we recommend the readers to discuss these two effects separately.

4. CONSTRAINTS ON NONSTANDARD EXPANSION PHASES

In this section, we will discuss the known constraints on extra components that can affect the expansion of the Universe, as well as some future ways to put such scenarios into test. We will discuss BBN and CMB constraints and observational prospects for detecting microhalos and primordial black holes, as well as stochastic gravitational wave backgrounds.

4.1. BBN and CMB constraints (Authors: M. Escudero & V. Poulin)

In this section we will discuss how observations of the CMB and the primordial element abundances can be used to obtain information about both the early and late (pre- and post-BBN) expansion history of the Universe, and how the topic has been discussed in the literature. To illustrate the constraining power of current cosmological observations, we outline in Table 2 the up to date CMB and BBN constraints on key parameters constraining nonstandard expansion histories.

We start by discussing the BBN constraints. BBN represents the earliest epoch of the Universe from which we have data. The agreement between the observed primordial abundances of helium and deuterium with the predicted ones is a highly nontrivial success of the standard cosmological model. Equivalently, BBN serves as a powerful probe of new physics and nonstandard cosmological expansion histories.

4.1.1. A brief review of standard BBN

Big Bang Nucleosynthesis occurred when the Universe was about three minutes old. During BBN ($1 \text{keV} \lesssim T \lesssim 10 \text{MeV}$), the Universe was radiation dominated and the only relevant species were electrons, positrons, photons, three neutrinos species, and a small number of protons and neutrons, $n_b/n_\gamma \sim 10^{-9}$. At temperatures $T \gtrsim 10 \text{MeV}$ all of

| Parameter | Description | Cosmological Data | Current Constraint |
|-----------|-------------|-------------------|--------------------|
| $N_{\text{eff}}$ | Number of effective relativistic neutrino species ($N_{\text{eff}}^{\text{SM}} = 3.045$) | Planck legacy | $N_{\text{eff}} = 2.99 \pm 0.34, \Delta N_{\text{eff}} < 0.30$ [33] |
| $T_{\text{reh}}$ | Reheating temperature | Planck 2015 | $T_{\text{reh}} > 4.7 \text{MeV}$ [28] ($\phi \to e^+ e^- / \gamma \gamma$) |
| $w$ | Constant EoS of dark energy | Planck+BAO+SN1a | $w = -1.028 \pm 0.062$ [33] |
| $w(a) = w_0 + (1 - a)w_a$ | Time-evolving EoS of dark energy | Planck+BAO+SN1a | $w_0 = -0.96_{-0.15}^{+0.19}, w_a = -0.29_{-0.08}^{+0.08}$ [33] |
| $z_{\text{MR}}$ | Redshift at matter-radiation equality | Planck | $z_{\text{MR}} = 3400 \pm 50$ [33] |
| $z_e$ | Redshift at decoupling | Planck | $z_e = 1089.9 \pm 0.5$ [33] |

**TABLE 2**

**Summary of current constraints on various parameters constraining nonstandard cosmological expansion histories. All bounds or intervals are given at 95% CL.**
these species were tightly coupled via electromagnetic and weak interactions. However, the rate of weak interactions ($\Gamma \sim G_F^2 T^5$) fell below the expansion rate $H$ at a temperature $T \sim (G_F^2 M_p)^{-1/3} \sim 1$ MeV. In particular, neutrino-electron interactions froze out at $T_{\nu e}^{\text{dec}} \sim 2$ MeV [638]. From then on, neutrinos evolved as a decoupled fluid while at $T_{\gamma} \sim m_e$ electrons and positrons started to annihilate into photons heating up the photon fluid with respect to the neutrino one, leading to temperature ratio $T_\gamma/T_\nu \simeq (11/4)^{1/3} \simeq 1.4$ [17]. Soon after neutrino decoupling, at $T_{\text{np}} \sim 0.7$ MeV, interactions interconverting protons to neutrons also froze out ($t_{\text{np}} \sim 1$ s). At temperatures $T \gtrsim 0.1$ MeV, interactions between bounded nuclei and the plasma were highly efficient, and light nuclides were kept in kinetic equilibrium. During this epoch, given the high temperature of the plasma and the small baryon-to-photon ratio, every bounded nuclei was rapidly destroyed by energetic species in the plasma. This situation changed dramatically when the temperature of the Universe dropped below $T_D \simeq 0.073$ MeV (corresponding to $t_D \simeq 241$ s). At this point, the Universe was cold enough for photons not to be capable of disassociating deuterium any more. Soon after this happened, a chain reaction was triggered and almost all free neutrinos in the plasma formed helium ($^4$He). In particular, the standard BBN leads to a Universe in which roughly $\sim 75\%$ of the baryonic energy density is stored in the form of hydrogen and where the $\sim 25\%$ remaining baryonic mass is primarily in the form of helium. In addition, small (but relevant) abundances of heavier nuclei remained: a fraction $10^{-5} - 10^{-3}$ of deuterium ($^2$H) and tritium ($^3$He), and only very small ($\lesssim 10^{-9}$) abundances of heavier elements such as $^7$Li and $^6$Li were synthesised.

Within the standard BBN, the only parameter needed to predict the light element abundances is the baryon energy density $\Omega_b h^2$, or interchangeably the baryon-to-photon number density ratio $n_b / n_{\gamma} \simeq 0.00237 \simeq (6.112 \times 10^{-10})$ [639]. The helium and deuterium primordial abundances are now measured with 1% precision [639]. Their observed values are fully compatible with those predicted within the standard BBN and point to $0.021 < \Omega_b h^2 < 0.024$ at 95% CL [10]. This range is in excellent agreement with the values inferred from CMB observations within the framework of $\Lambda$CDM [33]. Hence, this agreement represents a clear triumph of the standard cosmological model.

### 4.1.2. BBN constraints on nonstandard cosmologies

BBN has been widely used as a laboratory to constrain new physics (for reviews, see Refs. [258, 641, 642]). Here, we will review the constraints that can be derived from BBN on some of the most popular nonstandard expansion histories. In particular, we will discuss cosmologies with dark radiation, EMDE (low reheating temperature), unstable massive relics, light interacting species, and modified theories of gravity.

**Dark radiation:** Perhaps the single most studied nonstandard cosmology in the context of BBN is that featuring extra noninteracting massless species, i.e., dark radiation. Dark radiation is a generic prediction of many extensions of the SM, see e.g. Refs. [643-646], and its energy density $\rho_{\text{rad}}$ is typically parametrized by the number of effective relativistic neutrino species as relevant for CMB observations,

$$\Delta N_{\text{eff}} \equiv N_{\text{eff}} - N_{\text{eff}}^{\text{SM}}, \quad N_{\text{eff}} = \frac{8}{7} \left(\frac{11}{4}\right)^{4/3} \left(\frac{\rho_{\text{rad}} + \rho_{\gamma}}{\rho_{\gamma}}\right),$$

where $N_{\text{eff}}^{\text{SM}} = 3.045$ [635-637]. The main implication of $\Delta N_{\text{eff}} > 0$ during BBN is to enhance the expansion rate of the Universe, since $H \propto \sqrt{\rho}$. $\Delta N_{\text{eff}}$ is primarily constrained by the primordial helium abundance because it is extremely sensitive to the time at which deuterium starts to form and it is only logarithmically sensitive to the baryon energy density. Pioneering BBN constraints set $\Delta N_{\text{eff}} < 5$ [647], while for many years until the early 2010s, $\Delta N_{\text{eff}} < 1$ [15, 642] was quoted as a typical robust BBN bound. At present, and thanks to recent very precise measurements of the primordial helium abundance [648-650], state-of-the-art BBN analyses yield $\Delta N_{\text{eff}} < 0.33$ [634, 651] at 95% CL. This represents a very stringent constraint on many extensions of the SM. For example, it strongly disfavours light sterile neutrinos in thermal equilibrium during BBN [652-657], and it precludes the existence of any number of massless species that were once in thermal equilibrium with the SM plasma but decoupled at $T < 100$ MeV.

**EMDE/low reheating temperature:** As discussed in Sec. 2, several extensions of the SM of particle physics predict the existence of heavy and long-lived particles, which may cause an EMDE. We denote these by $\phi$. Massive species dominating the early Universe should be unstable, as they would have otherwise overwhelmed the Universe. BBN is strongly sensitive to this EMDE and subsequent decay of massive particles, and the requirement of successful BBN typically leads to a constraint on the lifetime of species dominating the Universe, $\tau_{\phi} \lesssim 0.1$ s. The exact description of the early Universe dominated by an unstable massive relic at $t \gtrsim 0.001$ s is far from trivial and there are only a handful of references dealing with this issue in detail [23-29]. In these scenarios, the unstable particle dominating the Universe will decay into different states (typically within the SM) and will reheat the Universe to a temperature $T_{\text{reh}} \simeq 0.7$ MeV $\sqrt{\tau_{\phi}}$, defined via $T_{\phi} \equiv 3 H(T_{\text{reh}})$. BBN represents a key probe to test these scenarios since observations are compatible with the standard RD picture of the early Universe. The precise implications for BBN of such type of scenarios depend upon the exact decay channels of the out-of-equilibrium decaying particle, that we shall denote by $\phi$.

10 Only the measured abundance of $^7$Li does not match the standard BBN prediction. This is the so-called Lithium problem [639, 640]. At present, there is no clear resolution to the problem, but the two most likely avenues to resolve it are: i) a stellar depletion mechanism, or ii) new physics capable of producing less $^7$Li than in the standard BBN. We note that nonstandard expansion histories do not seem likely of being capable of ameliorating or solving the problem.

11 The small difference between $N_{\text{eff}}^{\text{SM}}$ and 3 mainly arises from the fact that neutrino decoupling (at $T \sim 2$ MeV) took place soon before electrons became nonrelativistic ($T \sim m_e$) and there were some residual $e^+e^-$ annihilations into neutrinos.

12 Note that, as discussed in Sec. 2.3, particle decays cannot increase the temperature of the Universe [38].
The latest BBN analysis yields $T_{\text{reh}} > (4-5) \text{ MeV} \, (\phi \to \text{hadrons})$ and $T_{\text{reh}} > 1.8 \text{ MeV} \, (\phi \to e^+e^-/\gamma\gamma)$, both at 95% CL. In addition, scenarios where $\phi$ possess a direct decay mode to neutrinos should fulfill $T_{\text{reh}} > (1-4) \text{ MeV}$ [25] (for $\phi \to \bar{\nu}e/\gamma\gamma$) at 95% CL. The ranges reflect the dependency of the bound on the decay mode and mass of $\phi$. Note that scenarios with a low reheating temperature are among the few situations exhibiting $N_{\text{eff}} < N_{\text{eff}}^{\text{SM}}$. Since CMB observations are sensitive to $N_{\text{eff}}^\phi$ they can also be used to constrain $T_{\text{reh}}$. The latest CMB analysis, using Planck 2015 data, yields $T_{\text{reh}} > 4.7 \text{ MeV} \, (\phi \to e^+e^-)$ at 95% CL.

To summarize, since $T_{\text{reh}} \approx 2.2 \text{ MeV} \sqrt{0.1s/\tau_{\phi}}$ and current bounds set $T_{\text{reh}} \gtrsim 2 \text{ MeV}$, successful BBN bounds the lifetime of unstable relics dominating the early Universe to be $\tau_{\phi} \lesssim 0.1 \text{ s}$.

**Unstable massive species:** Species with small abundances at the time of BBN, $\rho/\rho_r \ll 1$, will not significantly impact the expansion of the Universe during BBN. Yet, they are subject to very stringent constraints from BBN [658–664]. BBN constraints on these species arise from the fact that the decay products of relics with $m > 5 \text{ MeV}$ can initiate electromagnetic or hadronic cascades (depending on their mass) which can alter the network of BBN reactions through e.g. photo-disassociation of already synthesized nuclei. Since the baryon-to-photon ratio $\eta \sim 10^{-9}$, the decay of even very small number densities of particles can still drastically alter the light element abundances. BBN bounds on unstable relics are used to constrain the abundance of a wide array of relics with lifetimes in the range $0.01 \text{ s} \lesssim \tau \lesssim 10^{12} \text{ s}$. For state-of-the-art work constraining generic long-lived decaying particles with BBN, see Ref. [665]. Examples of detailed BBN constraints on concrete and well-motivated particle physics scenarios include gravitinos [666], dark scalars [667, 668], dark vector bosons [669, 670], MeV-scale dark sectors [671–673], GeV-scale sterile neutrinos [674–676], and PBHs [677].

The take away message from these studies is that BBN represents a very stringent constraint on the abundance of unstable massive relics with lifetimes in the range $0.01 \text{ s} \lesssim \tau \lesssim 10^{12} \text{ s}$.

**Light interacting species:** BBN is sensitive to the Universe in the temperature window $1 \text{ keV} \lesssim T \lesssim 10 \text{ MeV}$. This makes BBN a powerful tool to test extensions of the SM featuring particles with masses at the MeV scale and below. Examples of such particles include axions, majorons, dark photons, and generic dark sector states. The consequences on BBN of light ($m \lesssim 10 \text{ MeV}$) BSM states depend upon the exact properties of the given particle and its abundance at the time of BBN. Light particles in thermal equilibrium with the SM plasma at $T \lesssim 10 \text{ MeV}$ will have an energy density comparable to that of the SM plasma and will therefore alter the expansion history of the Universe. In addition, these particles will interact with SM species and eventually, at $T \sim m/3$, will release entropy into the SM plasma via their decays or annihilations. The cosmological consequences of such scenarios were first highlighted in Ref. [678]. The most recent BBN analysis shows that generic species in thermal equilibrium during BBN should have a mass $m > 0.4 \text{ MeV}$ [679] at 95% CL, and that light species ($m < 0.1 \text{ MeV}$) in thermal equilibrium with the SM plasma are highly disfavoured by current cosmological observations. BBN represents a very stringent constraint on light species that were not in thermal equilibrium during BBN but that were still capable of modifying the expansion history of the early Universe. Relevant examples of these scenarios include axions [680, 681], axion-like particles [682, 683], majorons [684, 685], and weakly coupled dark sectors [686, 687].

**Modified gravity theories:** So far we have only discussed the impact on BBN of nonstandard cosmologies featuring BSM matter. However, modified gravity theories can also modify the expansion history of the early Universe. In this context, BBN has been used to constrain, for example, large extra dimensions [688, 689], the value of Newton’s constant during nucleosynthesis [690–692], and scalar-tensor theories of gravity [693–695].

**BBN tools:** We finish by pointing the reader to publicly available state-of-the-art tools used to model the early Universe and BBN: NUDEC_BSM [637, 696], PARThEnOPE [697, 698], AlterBBN [699, 700], and PRIMAT [651]. Given current and expected precision on BBN observables, these tools prove very useful in studying the impact of nonstandard cosmological expansion histories on BBN.

4.1.3. Effects of nonstandard expansion rate on the CMB

The tremendous increase in the precision of CMB observations over the last 20 years have led to a remarkably accurate determination of the $\Lambda$CDM parameters, some of which are nowadays known at precision better than 1%. This is in particular true for the angular scale of sound horizon $\theta_s(z_{\text{rec}})$, that has been determined at 0.04% with Planck [33]. The $\Lambda$CDM model teaches us that the Universe was dominated by radiation until redshift $z \sim 3500$, then by matter until $z \sim 0.7$, and finally by dark energy. Given such high accuracy measurements, it is natural to expect that strong constraints apply on deviations from the $\Lambda$CDM expansion history.

In practice, however, the constraining power of the CMB greatly changes before and after decoupling. In fact, it is only by combining CMB with measurements of the Baryonic Acoustic Oscillations (BAO) (see e.g. Refs. [704–707]) and Supernovae of type 1a (SN1a) (see e.g. Refs. [702]) at $z \lesssim 2$ that one can constrain modifications to $\Lambda$CDM to high accuracy in the late Universe. As of yet, the era from (roughly) $2 \lesssim z \lesssim 1000$ remains largely unconstrained. Future ground-based 21 cm surveys such as HERA [708] and SKA, and other intensity mapping techniques, will allow to extend constraints to the cosmic dawn epoch at $z \lesssim 20$ [709–713], while future Moon-based experiments could probe

13 https://www.skatelescope.org/
the dark ages until \( z \lesssim 200 \) \cite{714–716}. At \( z > 1000 \), on the other hand, modifications of the expansion history affect the CMB spectra in ways that are much less subject to degeneracies. Still, the CMB cannot probe the expansion history up to arbitrarily high \( z \). A multipole mode \( \ell \) can be approximately mapped onto a physical wavenumber \( k \) via \(^{14} \ell \sim k \eta_0 \) with \( \eta_0 \sim 10^5 \) Mpc the conformal time today. As a mode is mostly affected by an exotic expansion history if it is within the cosmological horizon, we expect that it will probe the expansion at times \( \eta \gtrsim \eta_0/\ell \). We show the minimum multipole affected by a modified expansion history at \( (\text{and below}) \) \( z \) in Fig. 10, left panel \cite{701} and illustrate the maximum redshift probed by \( \text{WMAP, Planck} \), and a futuristic experiment like \( \text{CMB-S4} \) (assuming that no too big modifications to the expansion history are allowed when relating \( \eta \) to \( z \)). Concretely, \( \text{Planck} \) is sensitive to exotic histories up to \( z_{\text{max}} \sim 80,000 \), while an experiment like \( \text{CMB-S4} \) would probe up to \( z_{\text{max}} \sim 160,000 \).

In the following, we review how CMB measurements can be used to learn about the expansion history of the Universe from \( z \sim 10^5 \) to today. We discuss “model-independent” modifications\(^{15} \) to \( H(z) \) at \( z \gg 1000 \) and \( z \ll 1000 \). In practice, however, details of perturbations are important when deriving constraints on exotic expansion history. We will illustrate these subtleties with two specific scenarios: exotic neutrino properties and early dark energy.

**Pre-decoupling era:** Well before decoupling, photons and baryons form a tightly coupled fluid. In the Newtonian gauge (see e.g. Ref. \cite{342} for definitions), super-horizon perturbations are frozen, while sub-horizon density perturbations \( \delta_x \equiv \delta \rho_x/\rho_x = 3/4 \delta_b \) experience driven acoustic oscillations whose dynamics is governed by \(^{16} \delta_x(x) = [\delta_x(0) + \Phi(0)] \cos(x) + 2 \int_0^x dx' \Phi(x') \sin(x - x') \) where \( \Phi \) is the Newtonian gravitational potential and \( x = kc, \eta \) \cite{701}.

Therefore, well before decoupling exotic expansion histories can mainly influence (i) the sound horizon \( r_s = c_s \eta \) and (ii) the time evolution of the gravitational potential. The most important scale in the CMB is the sound horizon at decoupling, which is the distance travelled by sound wave in the primordial plasma from reheating until decoupling, computed as (see e.g. Ref. \cite{17})

\[
 r_s(z_{\text{rec}}) = \frac{1}{1 + z_{\text{rec}}} \int_{z_{\text{rec}}}^\infty \frac{dz}{H(z)} c_s(z), \quad c_s(z) = \frac{c}{\sqrt{3(1 + R)}}, \tag{63}
\]

where \( c_s(z) \) is the sound speed in the tightly coupled photon-baryon fluid and \( R = 3 \rho_b/4 \rho_\gamma \). This scale dictates the typical size of inhomogeneous patches in CMB maps, and accordingly the position (or more accurately the spacing) of the acoustic peaks. Note that this scale is weighed towards smaller \( z \), and is therefore more sensitive to change to \( H(z) \) close to decoupling.

On the other hand, the decay of the gravitational potentials on sub-horizon scales during radiation domination lead to a boost in the amplitude of oscillations for modes \( k > k_{\text{MR}} \equiv H(z_{\text{MR}})/(1 + z_{\text{MR}}) \) \cite{718}. This, in turn, increases the amplitude of the acoustic peaks at \( \ell > \ell_{\text{MR}} \sim k_{\text{MR}} \eta_0 \). It is mostly through this effect that CMB measurements allow to put strong constraints on the redshift of matter-radiation equality \( z_{\text{MR}} \sim 3400 \pm 50 \) \cite{33}. Modifications to the standard time evolution of the gravitational potential at a given redshift \( z \) therefore lead to change in the amplitude of acoustic peaks around the mode \( \ell \sim k \eta_0 \), where \( k = H/(1 + z) \) is the scale that entered the horizon around \( z \). Interestingly, this scale-dependence can be used to trace the time at which the expansion rate is modified compared

\(^{14} \) In reality, a given \( \ell \) is related to \( k \) via \( \Theta(k, \eta_0) \equiv \Delta T/T|k \sim \int d\eta S(k, \eta) J_\ell(k(\eta - \eta_0)) \), where \( S(k, \tau) \) is the so-called “source function” which encodes various contributions to the temperature anisotropies and \( J_\ell \) is a Bessel function peaked around \( k \eta_0 \) (see e.g. Ref. \cite{717}). While several \( k \) modes can therefore contribute to a given \( \ell \), we make the approximation that the Bessel function can be described by a Dirac delta function for the sake of the discussion.

\(^{15} \) A detailed study of the sensitivity of the CMB to cosmological parameters can be found in e.g. Ref. \cite{17, 718}.

\(^{16} \) This solution is valid at length scales shorter than the sound horizon \( r_s \), neglects the time dependence of the sound speed of the baryon-photon fluid, any sources of anisotropic stress, and the damping terms due to baryon friction.
to $\Lambda$CDM.

Close to decoupling, the tight-coupling approximation breaks down and the photon mean free path $\lambda \equiv (n_e \sigma_T)^{-1}$ becomes significant. This leads to photon diffusion on a typical (comoving) distance $r_d^2(n_{\text{rec}}) \propto \int d\eta/\lambda$ (see e.g. Ref. [719] for a more accurate definition). This defines the scale $k_d = r_d^{-1}$ above which modes $\ell > k_d \eta_0$ are exponentially suppressed. Change in the expansion history close to decoupling can therefore affect the damping tail that is well constrained by the data. This effect is particularly important to constrain extra relativistic degrees of freedom (although partially degenerate with change in the spectral index $n_s$ or the primordial helium fraction $Y_p$) [720].

Finally, the decoupling process itself depends, in general, on the expansion history. However, the decoupling redshift $z_{\text{rec}}$ is only mildly affected by change in $H(z)$. This can be understood by noting that $z_{\text{rec}}$ can be approximately obtained by equating the Thomson interaction rate with the expansion rate $H(z) \sim x_e(z) n_H(z) \sigma_T$. Hence, a higher $H$ leads to a higher residual free-electron fraction $x_e \equiv n_e/n_H$ but does not change $z_{\text{rec}}$ to a first approximation.

**Post-decoupling era:** After decoupling, the main effect of changes in $H(z)$ is to change the distance that separates us from the last scattering surface, as known as the angular diameter distance to decoupling:

$$D_A(z_{\text{rec}}) = \frac{1}{1+z_{\text{rec}}} \int_0^{z_{\text{rec}}} \frac{dz}{H(z)},$$

(64)

or, equivalently, the conformal time today $t_0$. Once the pre-decoupling era is fixed, $D_A(z_{\text{rec}})$ (or $t_0$) determines the overall location of the spectrum in multipole space. A change in $D_A(z_{\text{rec}})$ will therefore shift the spectrum horizontally. It also dictates the multipole $\ell_d$ above which perturbations are damped due to diffusion and the multipole $\ell_{\text{eq}}$ above which scales have been gravitationally enhanced by the gravitational potential decay during RD era.

However, this pure geometric effect suffers strong degeneracy, since species with different $H(z)$ at late times can lead to the same $D_A(z)$. There is, fortunately, a number of additional effects that help in breaking the degeneracy within CMB data. In particular, the late-time evolution of potential wells manifests itself as a sudden rise in the large scale anisotropies. Because different multipoles are affected at different times, as discussed above, this gives the CMB spectrum some sensitivity to $H(z < z_{\text{rec}})$ (and not only to the integrated $H^{-1}(z \leq z_{\text{rec}})$). Unfortunately, however, at small $\ell$ the power spectrum $C_\ell$ suffers a large theoretical uncertainty known as cosmic variance, which scales as $\Delta C_\ell/C_\ell = \sqrt{\ell/(2(\ell+1))}$ and severely limits the ability of the CMB to measure $H(z < z_{\text{rec}})$.

4.1.4. Combining CMB with BAO and SN1a

The CMB is not the only cosmological probe at our disposal in the post-decoupling era, and one can therefore break the geometric degeneracy that exists in CMB data. The canonical examples of such probes are SN1a, powerful standard(ized) candles that can be used to measure the luminosity distances $D_L(z)$ up to Gpc scales ($z \lesssim 1$ currently), and have led to the discovery of modern accelerated expansion compatible with the Universe being dominated by a cosmological constant [721, 722]. Another important example is the BAO [723], which has been detected to high accuracy in the clustering of galaxies [724, 725], and can be used to measure $D_A(z)$ (currently until $z \lesssim 0.5$ [704]). More recently, the BAO has been observed in the (cross and auto-) correlation of quasar Ly-$\alpha$ forest, which allows to constrain the expansion history at $z \sim 2.4$ [705–707]. However, since we can only measure fluxes ($F \equiv L/4\pi D_L^2$) and angles ($\theta \equiv r_s(z_{\text{drag}})/D_A$), observations of SN1a and detection of the BAO do not allow to measure an absolute expansion rate; rather, by themselves these measurements constrain the shape of the expansion history. In practice, the CMB allows us to calibrate the BAO by providing a model-dependent measurement of the sound horizon at the redshift $z_{\text{drag}} \sim 200$ at which baryons decouple from photons $r_s(z_{\text{drag}})$.

The combination of CMB, BAO and (uncalibrated) SN1a data can be used to set constraints on a wide variety of dark energy and dark matter models which we do not attempt to review in detail here (a few canonical constraints are given in Table 2). The canonical example is the equation of state of DE, often parametrized as $w(a) = w_0 + (1-a)w_a$ [729] that must nowadays satisfy $w_0 = -0.961 \pm 0.077$ and $w_a = -0.28^{+0.31}_{-0.27}$ (68% CL) [33]. We show in the right panel of Fig. 10 the constraints on the EoS $w_0$ (assuming $w_a = 0$) and fractional matter density today $\Omega_m = 1-\Omega_A$ obtained from a combination of Planck data [33], the Pantheon SN1a catalogue [702] and the BOSS DR12 BAO measurement [704]. This clearly illustrates the geometric degeneracy that exists in CMB data: it is only by combining CMB with BAO and/or uncalibrated SN1a that one can unambiguously detect the presence of DE with an equation of state compatible with that of a cosmological constant. Another interesting example is that of DM decays on cosmological time-scales. A combination of CMB, BAO and LSS data sets a robust model-independent bound on the lifetime of DM, $\tau > 158$ Gy, and constrain the fraction of DM that has decayed between now and decoupling to be less than 4% [367, 733]. Alternatively, it is possible to obtain model-independent constraints on the expansion history at late times from CMB, galactic BAO and uncalibrated SN1a. These restrict deviation from $\Lambda$CDM to be less than 10% percent at $z \lesssim 1$ [734–736]. However, there is a number of interesting tensions between Planck and probes of the late-time expansion rate that might start indicating deviation from a cosmological constant [737].

17 Alternatively, it is possible to measure $r_s(z_{\text{drag}})$ by combining BAO and the value of $w_0$ obtained from BBN [726–728].

18 These bounds were derive assuming massless products. Decay into warm DM lead to stronger constraints because of the impact of the warm component on structure formation [730]. Electromagnetic decay products change the thermodynamics history and are much more strongly constrained [731, 732].
In particular, several methods have been proposed to calibrate the intrinsic luminosity of SN1a and obtain direct constraints on the Hubble constant \( H(z = 0) \equiv H_0 \) (see e.g. Refs. [703, 738]). In fact, the latest calibration of SN1a using cepheid stars [703] has led to a measurement of \( H_0 \) with 1.4% precision that is in 4.4σ (statistical) disagreement with what is predicted in the ΛCDM cosmology deduced from Planck CMB data [30]. In fig. 10, right panel, we show that because of the geometrical degeneracy discussed previously, one can easily accommodate such a value of \( H_0 \) by allowing for a “phantom” equation of state of dark energy \( w_0 < -1 \). However, the inclusion of BAO and Pantheon SN1a catalogue breaks that degeneracy, such that late-universe modifications as a resolution to the Hubble tension are tightly constrained [734, 736, 739, 740]. Barring the possible presence of systematics in the data (see e.g. Ref. [30] for a discussion), there is a growing consensus that it is necessary to decrease the value of the sound horizon compared to that deduced in the ΛCDM cosmology in order to resolve the tension rather than modifying the late-expansion history [734, 739, 740]. If confirmed, this discrepancy could therefore be a sign of an epoch of exotic expansion during the pre-decoupling era, as induced for instance by the presence of dark energy at early times [352, 353, 355, 359–361] or neutrinos with exotic properties [377, 741].

4.1.5. Probing perturbation dynamics with CMB

Until now, we have limited our discussion to modifications of the background expansion rate. However, the CMB is very sensitive to the detailed dynamics of perturbations in various fluids that contribute to the stress–energy tensor of the Universe. This is because, as argued previously, the time evolution of the gravitational potential (\( \Phi \)); previous section) leaves clear imprints on the CMB sky, and the evolution of the gravitational potential is related (via Einstein’s equations) to the sum of the perturbed stress-energy tensor of each fluid. Therefore, species with the same background evolution but different perturbation dynamics can have very different signatures in the CMB. Moreover, in a perturbed Universe, it is not consistent to neglect perturbations in a fluid\(^{19}\), as it violates the conservation equations that follow from Bianchi identities. We illustrate this important subtlety for two specific scenarios, exotic neutrinos and early dark energy.

**Neutrino properties:** As mentioned previously, the CMB can put strong constraints on neutrino properties, and more generally on any type of extra radiation density, often parametrized by \( \Delta N_{\text{eff}} \). However, bounds can strongly vary depending on the choice made for describing perturbations of the species, which are directly tied to the microphysical properties of neutrinos (or the extra species). The standard constraint on \( \Delta N_{\text{eff}} \) is obtained by assuming that the extra radiation can free-stream, similarly to SM neutrinos. A free-streaming species leads to a specific phase-shift in the position and to small change in the amplitude of the acoustic peaks in the CMB [742]. In fact, the “neutrino drag” effect produced by SM neutrinos has been claimed to be robustly detected for the first time in Planck data [743]. On the other hand, a combination of CMB and BAO data leads to \( \Delta N_{\text{eff}} < 0.30 \) at 95% CL. Interestingly, it is possible to completely change this constraint by relaxing the assumption of free-stream. In fact, CMB data are powerful in testing new interactions of neutrinos since these interactions affect the streaming properties of neutrinos.

As a striking example, it has been found that the CMB temperature data can be equally well fit by postulating that neutrinos strongly self-interact through a contact interaction [377, 378, 744, 745]. In this cosmology, the phase shift of the acoustic peaks is completely erased, but the position of the peaks is kept constant by increasing \( H_0 \). In fact, Planck 2015 temperature data\(^{20}\) can fully accommodate 4 species of strongly interacting neutrinos, with \( H_0 = 72.3 \pm 1.4 \text{ km/s/Mpc} \) [377]. Such model would fully resolve the Hubble tension, while having interesting consequences for laboratory experiments. Unfortunately, the simplest realization of this model is strongly constrained [746], but such degeneracy provide an interesting avenue to resolve the Hubble tension. In fact, scenarios with light mediators are substantially less constrained by laboratory data. These can also suppress neutrino free-streaming and therefore reduce the tension [685, 741, 747–749].

Looking forward, future CMB missions are expected to deliver high precision measurements of \( N_{\text{eff}} \). In particular, the upcoming Simons Observatory [750] is expected to constrain \( \Delta N_{\text{eff}} < 0.12 \) at 95% CL, and proposed experiments such as CMB-S4 [751] could constrain \( \Delta N_{\text{eff}} < 0.06 \) at 95% CL. These upcoming constraints could have a very important impact on many particle physics scenarios. In particular, if future measurements are in agreement with the SM expectation, CMB-S4 measurements could rule out massless dark radiation that decoupled from the SM plasma at temperatures above the QCD phase transition [33, 752].

**Early dark energy:** The possibility of the presence of dark energy before last-scattering has been studied for more than a decade [753, 754]. However, these models have recently become increasingly interesting in the context of the Hubble tension mentioned above. Indeed, it has been shown in several studies [352, 353, 355, 359–361] that this tension could indicate the presence of a frozen scalar field acting like dark energy until \( z \sim z_{\text{MR}} \), where it contributes to \( \sim 10\% \) of the total energy density of the Universe, and subsequently dilutes like radiation or faster. Strikingly, Planck data (high-\( \ell \) polarization in particular) not only constrain the background evolution of such field but also the dynamics of perturbations. Indeed, it has been found that the field must have an effective sound speed (in its rest frame) \( c_s^2 = 0.8 \pm 0.1 \), which can be achieved, for instance, through a noncanonical kinetic term [360] or an oscillating potential which flattens close to the initial field value [353, 355]. Models with different perturbation dynamics can only lead to a milder alleviation of the tension [352, 359, 361].

\(^{19}\) The exception is a cosmological constant which has an equation of state \( w = -1 \) and does not develop perturbations.

\(^{20}\) Note that the inclusion of Planck 2015 polarization data restricts the success of the solution.
4.2. **Observational probes of microhalos** (Authors: M. S. Delos & A. L. Erickcek)

4.2.1. **Gravitational signatures**

The only guaranteed signatures of dark matter halos are gravitational. Unfortunately, detecting the gravitational effects of the microhalos generated in nonstandard cosmologies is extremely challenging because microhalos are small and diffuse. Since deviations from radiation domination only affect the growth of subhorizon perturbations, the masses of these microhalos must be less than the dark matter mass contained within the horizon when the radiation temperature was 3 MeV, which implies a maximum initial microhalo mass of 1200 $M_\odot$ [43]. The radii of these microhalos depends on their formation time, but even the earliest-forming microhalos are still diffuse enough that their photometric microlensing signatures are far weaker than those generated by PBHs of the same mass [755, 756].

The small changes in the positions of background stars due to the motion of an intervening halo offer an alternative avenue for detection, but these astrometric microlensing signatures are only potentially observable for standard halos with masses greater than $10^3 M_\odot$ [757, 758]. Halos that form early in the matter-dominated era ($z \approx 1000$) from enhancements to the small-scale matter power spectrum are easier to detect because these ultracompact minihalos (UCMHs) [755] have denser central regions than later-forming halos. Even so, an astrometric lensing search by Gaia is only sensitive to UCMHs with masses greater than a solar mass [759]. Moreover, this analysis assumed that UCMHs have density profiles with $\rho \propto r^{-9/4}$ at small radii, as predicted by [755]. N-body simulations have revealed that this density profile is not realized in even the rarest and most isolated halos that form from Gaussian initial conditions [766, 767]. Instead, N-body simulations have widely shown that the first halos develop radial density profiles that scale as $\rho \propto r^{-3/2}$ at small radii [46, 767–772]. Moreover, successive mergers between halos with $\rho \propto r^{-3/2}$ inner density profiles gradually drive them toward shallower NFW density profiles [766, 773, 774]. Shallower density profiles generate weaker astrometric microlensing signatures [757], so sub-solar-mass halos are well beyond the reach of astrometric microlensing [775].

Recently, two other gravitational probes have been proposed that have the potential to detect sub-earth-mass halos: pulsar timing and stellar microlensing. A pulsar timing array can detect the presence of dark matter halos due to both the Shapiro time delay when light from a pulsar passes through a halo [776] and the Doppler effect when a halo accelerates either the pulsar or the Earth [777]. With twenty years of pulsar observations with SKA, the Doppler signal can be used to constrain the abundance of sub-earth-mass microhalos with NFW profiles and concentrations greater than 100 [756]. Although such high concentrations are not expected in standard cosmologies [778], the early formation of microhalos following an EMDE may yield such profiles [340]. Sub-earth-mass halos may also be detected by monitoring background stars as they are magnified by incluster stars near lensing caustics within an intervening cluster: the presence of microhalos within the cluster would induce detectable magnification variations [779]. While [779] only considered the signatures of halos formed from small-scale axion isocurvature perturbations, similarly dense and abundant halos can form from adiabatic initial conditions following an EMDE.

4.2.2. **Dark matter annihilation within microhalos**

If dark matter is produced thermally, then it will self-annihilate. These annihilations generally produce high-energy observable particles, and numerous attempts have been made to detect these indirect dark matter signatures [e.g. 780–783]. Since the annihilation rate is proportional to the square of dark matter density, the presence of structure boosts the annihilation rate. In standard cosmologies, the presence of structure enhances the cosmic annihilation rate by a factor of $\sim 10^9$ [784] and the presence of substructure enhances the annihilation rate within galaxies by a factor of $\sim 10$ [778]. Since the microhalos generated by nonstandard cosmologies are far denser and more abundant than microhalos in standard cosmologies, these boost factors are significantly higher in such scenarios: cosmic boosts following an EMDE can exceed $10^{16}$ [341] and boosts within dwarf galaxies can exceed $10^6$ [340]. These large boost factors can compensate for the reduced annihilation cross section required to generate the observed dark matter abundance when there is entropy production following dark matter freeze-out. Therefore, searches for signatures of dark matter annihilation offer a promising way to constrain nonstandard thermal histories.

The first attempts to use gamma-ray observations to constrain nonstandard cosmologies used Fermi-LAT constraints on dark matter annihilations within dwarf spheroidal galaxies (dSphs) [339, 340], but later analyses revealed that the isotropic gamma-ray background (IGRB) provides more powerful bounds on microhalo-dominated emission [341, 785]. Following an EMDE, most of the dark matter is bound into early-forming microhalos that track the dark matter density, and the dark matter annihilation rate within these dense microhalos far exceeds the annihilation rate outside these microhalos. If the central regions of these microhalos are not significantly affected by interactions with other structures, the annihilation rate per dark matter mass is homogeneous and constant after the microhalos form. Consequently, microhalo-dominated annihilation can be cast as an effective dark matter lifetime: $\tau_{\text{eff}} = \left[ 2 m_{\text{DM}} (\Gamma/M) \right]^{-1}$, where $m_{\text{DM}}$ is the mass of the dark matter particle and $\Gamma/M$ is the annihilation rate per dark matter mass. 

Observational lower bounds on $\tau_{\text{eff}}$ correspond to lower bounds on the dark matter lifetime for particles with mass $2 m_{\text{DM}}$ [341].

The annihilation rate per mass depends on the microhalos’ density profiles and their abundance. Ref. [604] showed

---

21 While halos with masses greater than a solar mass do not probe the expansion history between inflation and BBN, they provide valuable constraints on the primordial power spectrum generated during inflation [46, 759–765].
that microhalo profiles and their abundance can be predicted in a simple way from properties of the precursor linear density field, enabling rapid computation of $\Gamma/M$ in a given cosmological scenario [785]. Other works [339–341] have employed Press-Schechter theory [605] to characterize microhalo populations and assumed that these microhalos have NFW profiles with low concentrations at formation. The two approaches yield $\Gamma/M$ values that differ by a factor of $\sim 3$ [785], but in both approaches, $\Gamma/M$ is strongly dependent on the formation time of the first microhalos (as determined by the ratio of the cut-off scale to the reheating horizon scale) and largely insensitive to the reheat temperature.

Figure 11 illustrates how gamma-ray observations can constrain thermal dark matter production in nonstandard cosmologies. This figure shows constraints on the dark matter velocity-averaged annihilation cross section and the dark matter particle mass for an EMDE with a reheat temperature of 10 MeV. The shaded regions cannot generate the observed dark matter abundance, and the blue line shows the annihilation cross section required to generate the observed dark matter abundance in a standard cosmology. All points in the white region can yield the observed dark matter abundance if the EMDE is preceded by a radiation-dominated era [780]. The black dashed line shows limits on dark matter annihilation given standard structure formation [780], while the green solid line shows how microhalos’ contribution to the IGRB can enhance these constraints [785]. Ref. [341] demonstrated how observations of the IGRB severely restrict hidden-sector dark matter models that include EMDEs, provided that these EMDEs are too short to permit structure formation prior to reheating.

Finally, we note that early-forming microhalos would enhance the dark matter annihilation rate prior to reionization, when the energy injection from dark matter annihilations can affect CMB temperature and polarization anisotropies and the temperature of the intergalactic medium (IGM). The CMB already establishes powerful limits on dark matter annihilation, see e.g. Ref. [788]. For standard models of structure formation, dark matter halos form too late to significantly impact these constraints [789], but if most of the dark matter is bound into microhalos at $z \gtrsim 800$, the CMB could provide a new avenue for their detection. Similarly, dark matter annihilations within microhalos can dramatically increase the IGM temperature, leading to 21cm emission (as opposed to absorption) at $z \gtrsim 50$, long before conventional astrophysical sources of high-energy photons are expected to exist [790].

4.2.3. The microhalo population within galaxies

Since the microhalos’ gravitational signatures and annihilation signatures both depend on the microhalos’ density profiles and abundance, we must understand how subsequent structure formation affects the microhalo population. Most microhalos eventually become subhalos within much larger, later-forming halos, such as those that surround galaxies. Within the deep potential wells of galactic halos, microhalos are subjected both to tidal forces and to energetic encounters with other objects.\(^{22}\) The vast difference in scales between these microhalos and their hosts precludes direct numerical simulation of this picture, but a variety of analytic and semianalytic treatments have been developed.

To predict the impact of tidal forces, the simplest and most widely used method involves the tidal radius, the radius within the subhalo beyond which the host’s tidal forces exceed the subhalo’s self-gravity. Any subhalo material lying beyond the tidal radius is expected to be stripped [for a review of tidal radius definitions, see 792]. A refinement to this

\(^{22}\) Subhalos are also subject to dynamical friction [791], which decelerates them. However, as this deceleration is proportional to the mass of the subhalo, it is expected to be negligible for microhalos.
approach involves conditioning tidal stripping on the energy of subhalo material instead of its radius [793]. Beyond this tidal truncation picture, material below the tidal radius is also heated by tidal forces, and the subhalo undergoes internal dynamics in response. No simple treatments exist for these effects, and models that predict tidal evolution are tuned to match numerical simulations. These models include [794], [795], [796], [797], [798], [799], and [800], which describe the time evolution of subhalos; and [801], [802], and [803], which describe how their density profiles respond to this evolution. The impact of tidal effects on microhalos depends strongly on the microhalos’ characteristic internal density, see e.g. Ref. [785], but as long as they possess divergent central density, tidal evolution is expected to be a continuous process that never fully destroys subhalos [804].

In addition to tidal evolution, microhalos inside larger structures are also subject to encounters with other orbiting objects, such as stars or other microhalos. Owing to the high virial velocities within galactic halos, these encounters are expected to be impulsive; that is, the encounter timescale is much shorter than the timescale for the microhalo’s internal dynamics. In this limit, the energy injected by the encounter may be computed readily, see e.g. Refs. [805–807]. Additionally, a general argument suggests that the microhalo’s density profile should relax to $\rho \propto r^{-4}$ at large radii [808]. However, the microhalo’s full response to the impulse remains complicated. The response of a collisionless system to an impulsive shock has been widely studied in N-body simulations, originally in the context of galaxies, see e.g. Refs. [809, 810] and star clusters [811–813] and later in the context of microhalos themselves [814–817]. Ref. [785] found that for microhalos produced by an EMDE, the impact of encounters with other microhalos is subdominant to that of encounters with stars. Due to the relative rarity of stars, the impact of stellar encounters is highly stochastic, with the few closest encounters dominating a microhalo’s evolution [785]. Additionally, unlike tidal evolution, a penetrative encounter can disrupt a microhalo’s divergent central density cusp [817], potentially leading to its destruction.

Finally, we remark that the impact of these evolutionary effects can discriminate between microhalo-dominated annihilation and dark matter decay. Microhalos near the centers of galactic systems suffer greater disruption compared to other microhalos, giving $\Gamma/M$ spatial dependence. Focusing on the Draco dwarf galaxy, [785] found that the gamma-ray signal morphology from annihilation within microhalos can differ significantly from that of dark matter decay because the former is suppressed by tidal evolution and encounters with stars. This discrepancy would be even more pronounced when comparing different systems. For instance, a comparison between the Galactic Center’s high stellar density would greatly suppress any signal from annihilation within microhalos therein, whereas the isotropic gamma-ray background suffers minimal suppression. Consequently, a comparison between the gamma-ray signals from both systems could distinguish between microhalo-dominated dark matter annihilation and dark matter decay. Since the efficiency of both tidal evolution and impulsive encounters is highly sensitive to the internal density of the microhalos, such comparisons could even serve to measure detailed microhalo properties.

4.3. Observational probes on primordial black holes (Authors: T. Harada & K. Kohri)

In this section we briefly review the current status of observational constraints on the abundance of PBHs. We also discuss how the curvature power spectrum can be constrained by using the PBH constraints. We begin by discussing PBH evaporation and its consequences.

4.3.1. PBH evaporation and its consequences

By emitting Hawking radiation [818] a PBH can evaporate. Their lifetime is given by

$$\tau_{\text{PBH}} \sim \frac{m_{\text{PBH}}^3}{M_p^3} \sim t_0 \left( \frac{m_{\text{PBH}}}{5.1 \times 10^{14} \text{ g}} \right)^3 \sim 10^{64} \text{ yrs} \left( \frac{m_{\text{PBH}}}{1 \text{ M}_\odot} \right)^3,$$

(65)

where $t_0 = 13.8$ Gyrs is the current age of the Universe [33] and $M_\odot \simeq 2 \times 10^{33}$ g denotes Solar mass. According to the Hawking process, a black hole has a finite temperature,

$$T_{\text{PBH}} \sim \frac{M_\odot}{m_{\text{PBH}}} \sim 0.1 \text{ MeV} \left( \frac{m_{\text{PBH}}}{10^{15} \text{ g}} \right)^{-1} \sim 10^{-9} \text{ K} \left( \frac{m_{\text{PBH}}}{1 \text{ M}_\odot} \right)^{-1},$$

(66)

which means that we may observe $\gamma$-rays or cosmic rays (electron, positron, neutrino, etc.) produced by a presently evaporating PBH. Thus, one can severely constrain the abundance of PBHs with masses around $10^{15}$ g by the non-detection of $\gamma$-rays or cosmic-ray positrons.

Due to Hawking evaporation, PBHs with masses smaller than $5.1 \times 10^{14}$ g should have evaporated by the current age of the Universe. Therefore, such PBHs cannot constitute DM. An unstable PBH whose mass is smaller than $\sim 10^9$ g (with lifetime being $\mathcal{O}(1)$ s) cannot be constrained by observations as strongly even by adopting the constraint from BBN which has a good sensitivity at around $m_{\text{PBH}} \gtrsim 10^9$ g, see Refs. [93, 677] and the discussion below.\textsuperscript{23} In the SM of particle physics, to maintain stability of the Higgs field at its vacuum expectation value, emission of high-energy particles from evaporating PBHs is dangerous. This can be used to constrain the evaporating PBHs even for $m_{\text{PBH}} < 10^9$ g. However, a concrete bound on the number density of PBHs has not been obtained [819].

\textsuperscript{23} In the SM of particle physics, to maintain stability of the Higgs field at its vacuum expectation value, emission of high-energy particles from evaporating PBHs is dangerous. This can be used to constrain the evaporating PBHs even for $m_{\text{PBH}} < 10^9$ g. However, a concrete bound on the number density of PBHs has not been obtained [819].
principle, however, evaporating PBHs can be constrained by production of hypothetical massive DM relics, such as the lightest supersymmetric particle (LSP) or Planck mass relics at which the evaporation terminates. From a nonthermal production of the LSP due to the evaporation, one can obtain an upper bound on $\beta$ in order not to exceed the observed DM density \[820-822\]:

$$
\beta \lesssim 10^{-18} \left( \frac{m_{\text{PBH}}}{10^{11} \text{ g}} \right)^{-1/2} \left( \frac{m_{\text{LSP}}}{100 \text{ GeV}} \right)^{-1} \left( \frac{m_{\text{PBH}}}{10^{11} \text{ g}} \right) \left( \frac{m_{\text{LSP}}}{100 \text{ GeV}} \right)^{-1},
$$

(67)

with $m_{\text{LSP}}$ being the mass of the LSP of the order of the weak scale. Quite recently, it was reported that this bound can be strengthened for lighter DM by considering free-streaming and an erase of small-scale fluctuations due to its initial momentum \[823, 824\]. Additionally, from a possible production of Planck mass relics at the last stage of the evaporation process, one can obtain an upper bound on $\beta$ in order not to exceed the DM density:

$$
\beta \lesssim 10^{20} \left( \frac{m_{\text{PBH}}}{g} \right)^{3/2},
$$

(68)

which applies for $1 \text{ g} \lesssim m_{\text{PBH}} \lesssim 10^7 \text{ g} \ [825]$. The above bounds depend on the detailed nature of DM, which is currently unknown. Therefore, if one omits these bounds due to such uncertainty, PBHs can be allowed to even dominate the early Universe for a short amount of time before they completely vanish away. Other possible probes of light PBHs are provided by GWs, that can be produced by nonlinear effects related to the large curvature perturbation at small scales which produced the PBHs \[826\], through Hawking radiation \[826-829\] or by a coalescence of binary PBHs before the evaporation \[827, 828, 830, 831\]. We also note that in some models, baryogenesis through nonthermal leptogenesis may be realized successfully by right-handed neutrinos produced by the huge amount of evaporating PBHs \[832\].

4.3.2. Observational constraints on PBHs

PBHs can naturally form in the early Universe by large density/curvature fluctuations generated during inflation or in other scenarios. To obtain a theoretical prediction for the abundance of PBHs, the physics of PBH formation must be understood. This gives a theoretical basis to use the observational data about PBHs to constrain the specific cosmological scenarios. In Fig. 12 we plot a summary of observational constraints on the abundance of PBHs as a function of the PBH mass. The $y$-axis is the fraction of the mass density of PBHs to that of CDM, i.e., $f = \Omega_{\text{PBH}}/\Omega_{\text{CDM}}$. The curves in this figure are from Ref. \[93\]. The reason we did not adopt the constraints coming from 1) femtolensing, 2) neutron stars, and 3) white dwarfs are also discussed in Ref. \[93\]. The quantities $\beta$, $f = \Omega_{\text{PBH}}/\Omega_{\text{CDM}}$, and the curvature perturbation power spectrum $P_{\zeta}$ are related to each other in the following way:

$$
f = \frac{\Omega_{\text{PBH}}}{\Omega_{\text{CDM}}} \sim \left( \frac{\beta}{10^{-18}} \right) \left( \frac{m_{\text{PBH}}}{10^{15} \text{ g}} \right)^{-1/2} \sim \left( \frac{\beta}{10^{-8}} \right) \left( \frac{m_{\text{PBH}}}{30 M_\odot} \right)^{-1/2} \sim 10^8 \left( \frac{m_{\text{PBH}}}{30 M_\odot} \right)^{-1/2} \sqrt{P_{\zeta} \exp \left[ \frac{-1}{18 P_{\zeta}} \right]}.
$$

(69)

Here we have assumed radiation had fully dominated after the end of the primordial inflation. By this relation (see Ref. \[93\] for more precise relations), one can transfer the bounds on $\beta$ to the bounds on $f$ or $P_{\zeta}$. Next, we introduce each constraint shown in Fig. 12 on the abundance of PBHs.

The red region excludes the evaporating PBHs through the following observations: One can constrain the abundance of PBHs through observations of the gamma-ray background of isotropic components \[677\] (see also Refs. \[854, 855\]), galactic components \[677, 856\], line emissions from annihilating electron-positron pairs \[833, 857, 858\], and cosmic-ray positrons by Voyager 1 \[834\]. These observations have excluded the possibility for PBHs to be the majority of DM for the masses of $10^{14.5} \text{ g} \lesssim m_{\text{PBH}} \lesssim 10^{17} \text{ g}$. The evaporating PBHs emit high-energy particles whose electromagnetic energy can modify the recombination history and background temperature. This potentially affects the observations of CMB anisotropy \[677, 732, 859-862\] and cosmological 21 cm line signals \[677, 835, 863\]. By the non-detection of those features one can constrain the abundance of PBHs for the masses of $10^{13.5} \text{ g} \lesssim m_{\text{PBH}} \lesssim 10^{14.5} \text{ g}$.

Emissions of high-energy hadrons, gamma-rays, and charged leptons are dangerous for light elements which are produced in BBN. That is because they induce hadronic and electromagnetic showers through subsequent scatterings off the background particles, and then the produced daughter particles can destroy light elements through inelastic hadron scattering and photodissociation \[661\]. Comparing theoretical predictions of BBN with observational abundance of light elements, one can exclude the masses of $10^9 \text{ g} \lesssim m_{\text{PBH}} \lesssim 10^{13.5} \text{ g}$ as a considerable fraction of DM \[677\] (see also Refs. \[732, 862\] for possible modifications only on photodissociation).

Nowadays there is strong motivation to study PBHs because the rate of GW detected by aLIGO \[864\] can be fitted by the event rate of coalescence of binary PBHs with masses $\sim 30 M_\odot \ [865]$ for the fraction $f \sim \mathcal{O}(10^{-3})$ (see also Refs. \[866-868\]). In other words, one can simultaneously constrain the number density of PBHs by observing GWs produced by binary PBHs with arbitrary masses. The two green regions are excluded due to the non-detection
Fig. 12.— Summary of observational constraints on the abundance of PBHs as a function of the PBH mass. The $y$-axis is the fraction of PBHs normalized to the energy density of CDM. The red region excludes the evaporating PBHs through the observations of gamma-ray background [677], annihilating electron-positron pairs [833], positrons by Voyager 1 [834], light element abundances with BBN [677], cosmological 21 cm signals [677, 835], and CMB anisotropy [677]. The magenta region excludes PBHs through gravitational lensing observations by HSC [836, 837], OGLE [838], EROS [839], MACHO [840] and Icarus [841]. From the observations of pulsar timing, one can constrain PBHs by the non-detection of GWs nonlinearly-produced by curvature perturbation [93, 842, 843]. Through the distortion of the global spectrum of CMB observations, one can obtain an upper bound on the curvature perturbation and the abundance of PBHs [93, 844, 845]. The green regions exclude PBHs by the non-detection of GWs by LIGO/Virgo [846, 847]. Because accretion of gas into a PBH can emit radiation at radio and X-ray frequencies, observational data exclude PBHs masses $\lesssim 10^3 M_\odot$ [848]. The cosmic accretion into PBHs is severely constrained by the observations of CMB polarization [849, 850]. The dissipation of curvature perturbation induces dilution of baryon-to-photon ratio which affects BBN [851–853]. The lines in this figure are from Ref. [93].

of such GWs by the current sensitivities of aLIGO/aVirgo for $0.2 M_\odot \lesssim m_{PBH} \lesssim 1 M_\odot$ and $10 M_\odot \lesssim m_{PBH} \lesssim 3 \times 10^2 M_\odot$ [846, 847] (see also Refs. [869, 870]). Future detectability of GWs from lighter binary PBHs and its consequences on a bound on PBH abundance are discussed in Ref. [871].

Observations of gravitational lensing can potentially give a constraint on an object passing through the line of sight towards a light source. The magenta region is excluded through gravitational lensing by the observations of HSC [836, 837], OGLE [838], EROS [839], MACHO [840] and Icarus [841]. From Fig. 12, we see that the masses of $10^{-10} M_\odot \lesssim m_{PBH} \lesssim 2 \times 10^4 M_\odot$ and $6 \times 10^4 M_\odot \lesssim m_{PBH} \lesssim 10^{9.5} M_\odot$ are excluded by the non-detection of lensing events caused by PBHs.

As discussed below, upper bounds on curvature perturbation $P_\zeta$ can be transformed into the ones on the abundance of PBHs ($f$ or $\beta$). That is because PBHs are expected to form in the early Universe by the collapse of inhomogeneous regions, which are characterized by the curvature perturbation. The dissipation of density perturbation at small scales in the early Universe induces dilution of baryons which affects the freeze-out value of neutron-to-proton ratio $n/p$ during BBN. Then, the brown region is excluded by comparing the theoretical predictions of light element abundances with the observations for $10^{4.5} M_\odot \lesssim m_{PBH} \lesssim 10^{5.5} M_\odot$ [851–853]. Through observations of the CMB $\mu$-distortion caused by the dissipation of density perturbation, one can obtain an upper bound on curvature perturbation and, simultaneously, the abundance of PBHs for $10^{4.5} M_\odot \lesssim m_{PBH} \lesssim 10^{13.5} M_\odot$ [93, 844, 845]. From the observations of pulsar timing, one can constrain PBHs by the non-detection of the GW background secondarily-induced by curvature perturbation [93, 842, 843, 872, 873].

Radio waves and X-rays are emitted through accretion of gas into a PBH. From observations, one can obtain an upper bound on the abundance of PBHs with masses $\mathcal{O}(10) - \mathcal{O}(10^2) M_\odot$, which is denoted by the cyan region [848]. Cosmological accretion is a relatively new subject. Considering a finite relative velocity between baryon and CDM/PBHs, disk accretion of baryons into PBHs must have occurred inevitably at high redshifts from $z \sim \mathcal{O}(10^2)$ to $z \sim \mathcal{O}(10^3)$. The photons emitted through the accretion can change the recombination history of baryons, which is severely con-
strained by observations of the CMB anisotropy and polarization. The yellow ocher region is excluded by the current observations of accretion into a PBHs and their CDM halo systems [849, 850] (see also Refs. [874–876]).

4.3.3. Constraints on the power spectrum of curvature perturbation

In earlier sections, we have discussed inflation and curvaton models which produce curvature perturbations. Actually, so far a lot of inflationary models have been proposed in which PBHs are produced at small scales while simultaneously fitting the CMB normalization at large scales. For example, there are models of double inflation [845, 877], hilltop inflation [878–880], preheating [188, 881], curvaton [204, 882], thermal inflation [233], false vacuum bubbles [883, 884], Q-ball formation [91, 885], and so on. These models predict that curvature perturbation can be $\mathcal{P}_\zeta \sim \mathcal{O}(0.1)$ at small scales due to special features in the spectrum, e.g., a blue-tilted spectrum, a sharp peak structure, and so on.

In particular, motivated by the blue-tilted spectrum predicted by some of the inflation models above, as an example, here we introduce a simple parametrization of large curvature perturbation at small scales as follows:

$$\mathcal{P}_\zeta(k) = \mathcal{A} \left( \frac{k}{k_*} \right)^{n_s-1} \frac{2}{\ln(\xi)} \left( \frac{\ln(\xi)}{\xi} \right)^2 + \cdots,$$

where $\mathcal{A}$ is the normalization of the amplitude, called the “CMB normalization” [886] at the pivot scale of the wavenumber $k = k_* = 0.05 \text{ Mpc}^{-1}$, $n_s$ is the spectral index, $\alpha_s$ is its running, and $\beta_s$ is the running of the running defined as

$$n_s - 1 = \frac{d \ln \mathcal{P}_\zeta}{d \ln k}, \quad \alpha_s = \frac{d n_s}{d \ln k}, \quad \beta_s = \frac{d^2 n_s}{d (\ln k)^2}.$$

This kind of parametrization considering the higher-order corrections up to $\beta_s$ was used by the Planck collaboration [887]. It also extends the definition used in Sec. 3.3. The Planck 2018 observations (with the TT, TE, EE+lowE+lensing dataset) give

$$n_s = 0.9625 \pm 0.0048,$$
$$\alpha_s = 0.002 \pm 0.010,$$
$$\beta_s = 0.010 \pm 0.013,$$

with $\ln(10^{10} \mathcal{A}) = 3.045 \pm 0.0016$ at 68% CL [13, 33]. Actually some inflation models predict such a blue-tilted spectrum with these quantities, see e.g., Refs. [878–880].

In Fig. 13, we plot the constraints on the curvature power spectrum $\mathcal{P}_\zeta$ as a function of wavenumber $k$. Here we have a simple relation between the power spectrum of the curvature perturbation $\mathcal{P}_\zeta(k)$ and the density perturbation, $\mathcal{P}_\zeta(k) \sim \sigma^2(5 + 3\omega)^2/2(1 + \omega)^2$ with $\sigma^2$ being the mean variance of density perturbation [888] (see also Ref. [624] for a more accurate relation). Allowed within 95% CL, we can see that the predicted curve of curvature perturbation with $n_s = 0.96$, $\alpha_s = 0.010$, and $\beta_s = 0.020$ can reach the edge of the bound from PBHs with the e-folding number $N \approx 17$. This means that a PBH with the mass $30 \text{ M}_\odot$, which can fit the LIGO events, can be produced in this parametrization [888] at $k = k_{\text{LIGO}} \sim 10^3 \text{Mpc}^{-1}$. See also Ref. [889] for similar analyses by using only $n_s$ and Refs. [89, 90] by using only $n_s$ and $\alpha_s$. Possible uncertainties about this kind of parametrizations are also discussed in Ref. [890]. Here we used the relation between the wavenumber $k = k_{\text{RD}}$, at which the mode entered into the horizon when the PBH was produced, and the mass of the PBH

$$k_{\text{RD}} \sim 6 \times 10^{15} \text{Mpc}^{-1} \left( \frac{m_{\text{PBH}}}{10^{15} \text{g}} \right)^{-1/2}.$$

4.3.4. Constraints on curvature perturbations from PBH formation during matter dominance

Finally, we discuss how one can constrain the abundance of PBHs produced during an EMDE. An EMDE may end by the decay of a massive particle $X$ such a the inflaton or curvaton field which had dominated the Universe and its subsequent thermalization. This transition is characterized by the reheating temperature $T_{\text{reh}}$ defined by $\Gamma_X \equiv 3H(T = T_{\text{reh}})$ with $\Gamma_X$ being the decay rate of the massive particle $X$ (see e.g., Sec. 2.3 for details). The wavenumber which crosses the horizon at $T = T_{\text{reh}}$ is defined by

$$k_R \equiv aH(T = T_{\text{reh}}).$$

If a PBH was produced during an EMDE, the relation between the energy fraction $\beta = \beta_{\text{MD}}$ at the time of formation time and the energy fraction in terms of CDM at present, $f$, is modified to [888, 892]

$$\beta_{\text{MD}} \sim 6 \times 10^{-18} f \left( \frac{T_{\text{reh}}}{10^{10} \text{GeV}} \right)^{-1}.$$
Fig. 13.— Constraints on curvature perturbation as a function of wavenumber (comoving momentum) \( k \) in units of \( \text{Mpc}^{-1} \). The red region is excluded by the constraints on PBHs [93]. The dashed curve denotes the upper bound in case of nonstandard scenarios, cf. Eqs. (67) and (68). From the CMB observations one obtains the allowed region at the pivot scale \( k = k_\text{piv} \approx 0.05 \text{Mpc}^{-1} \) [33, 891]. The green, orange, and brown regions are excluded by the \( \mu \)-distortion of the CMB spectrum [93, 844, 845], secondary GWs [93, 842, 843, 872], and the dilution of baryons (or the \( n/p \) ratio) after BBN [851–853], respectively. The lines in the figure are from Ref. [93].

After reheating, the standard RD epoch was realized. According to Eq. (69), we see that the abundance of PBHs produced during the RD epoch is

\[
\beta_{\text{RD}} \sim 10^{-18} f \left( \frac{m_{\text{PBH}}}{10^{15} \text{g}} \right)^{1/2},
\]

In addition, the relation between \( k \) and \( m_{\text{PBH}} \) for PBHs produced during a MD epoch is different from Eq. (73), as now [888, 892]

\[
k_{\text{MD}} \sim 2 \times 10^{16} \text{Mpc}^{-1} \left( \frac{m_{\text{PBH}}}{10^{15} \text{g}} \right)^{-1/3} \left( \frac{T_{\text{reh}}}{10^{10} \text{GeV}} \right)^{1/3}.
\]

Then, we can use the wavenumber

\[
k = \begin{cases} 
  k_{\text{MD}}, & \text{for } k \gtrsim k_{\text{R}} \\
  k_{\text{RD}}, & \text{for } k \lesssim k_{\text{R}}
\end{cases}
\]

and the abundance

\[
\beta = \begin{cases} 
  \beta_{\text{MD}}, & \text{for } k \gtrsim k_{\text{R}} \\
  \beta_{\text{RD}}, & \text{for } k \lesssim k_{\text{R}}
\end{cases}
\]

In Fig. 14, the constraints on the power spectrum of the curvature perturbation \( P_\zeta \) are plotted, as an example, without the spin suppression discussed in Sec. 3.6. The figure is from Ref. [89], where the authors adopted the relation \( \beta = 0.056 \sigma^2 \). The constraint is highly dependent on the reheating temperature, obeying a simple scaling \( P_\zeta \propto \sigma^2 \propto \beta_{\text{MD}}^{2/5} \propto T_{\text{reh}}^{-2/5} \). We stress the importance of the effects of the exponential suppression of \( \beta \) as a function of \( \sigma \) due to a finite spin (see Sec. 3.6). In Fig. 15, we show the constraints on \( P_\zeta \) depending on the reheating temperature, \( T_{\text{reh}} = 10^4 \text{GeV} \) (left panel) and \( T_{\text{reh}} = 10^9 \text{GeV} \) (right panel) by adopting the relation between \( \beta \) and \( \sigma \) for PBHs produced during an EMDE. Here we have used the result for the second-order effect shown in Fig. 8. As noted in Sec. 3.6.2, we recommend using this curve when looking for allowed regions of parameters in theoretical models which produce PBHs during an EMDE because it fits data more conservatively with milder assumptions of the theoretical model.
4.4. Gravitational wave backgrounds (Authors: D. G. Figueroa & M. Lewicki)

In this section we discuss three potential cosmological contribution to the stochastic GW background (SGWB), inflation, cosmic strings and phase transitions, and how a nonstandard expansion phase can effect the GW spectrum from these sources.

4.4.1. Inflation

During the inflationary period, tensor perturbations, originated out of quantum fluctuations, are spatially stretched to scales exponentially larger than the inflationary Hubble radius. This results in a (quasi-)scale invariant tensor power spectrum at super-horizon scales [893–896],

$$\langle h_{ij}(t, x) h^{ij}(t, x) \rangle \equiv \int \frac{dk}{k^3} \Delta^2_{h, \text{inf}}(t, k) \quad \leftrightarrow \quad \Delta^2_{h, \text{inf}}(k) \simeq \frac{2}{\pi^2} \left( \frac{H_{\text{inf}}}{m_{p}} \right)^2 \left( \frac{k}{k_*} \right)^{n_s} ; \quad n_t \simeq -2\epsilon \simeq -\frac{r_*}{8} , \quad (80)$$

where $k_*$ is a pivot scale and $H_{\text{inf}}$ is the inflationary Hubble scale when the mode $k_*$ crossed the Hubble radius during inflation. Slow-roll inflation leads typically to a slightly red “slow-roll suppressed” tilt, which can be expressed as proportional to the tensor-to-scalar ratio $r_*$ (evaluated at $k_*$). As the latter is constrained (at $k_*/a_0 = 0.002 \text{ Mpc}^{-1}$) by the B-mode in the CMB as $r_* \leq 0.06$ [13, 460], this leads to an upper bound on the scale of inflation $H_{\text{inf}} \lesssim H_{\text{max}} \simeq 6.6 \times 10^{13} \text{ GeV}$, implying also that the red tilt can only be very small $-n_t \leq 0.008 \ll 1$.

Following inflation the tensor modes re-enter successively back inside the horizon. After crossing the horizon, each tensor mode starts oscillating with decreasing amplitude $h_{ij} \propto 1/a$, independently of the expansion rate [897]. Since different modes re-enter the horizon at different moments of the cosmic history, the may propagate through different periods of the evolution of the Universe after becoming sub-horizon. In order to characterize the spectrum of the
SGWB today, we introduce the energy density spectrum, normalized to the critical density $\rho_{\text{crit}} = 3 m_p^2 H^2$, as [897]

$$\Omega_{\text{GW}}(t, k) \equiv \frac{1}{\rho_{\text{crit}}} \frac{d\rho_{\text{GW}}(t, k)}{d \ln k} = \frac{k^2}{12 a^2(t) H^2(t)} \Delta_k^2(t, k),$$

$$\Delta_k^2(t, k) \equiv T_h(t, k) \Delta_{h, \text{inf}}^2(k), \quad T_h(t, k) \equiv \frac{1}{2} \left( \frac{a_k}{a(t)} \right)^2,$$  \hspace{1cm} (81)

where $T_h(t, k)$ is a transfer function [898] that characterizes the expansion history between the moment of horizon re-entry $t = t_k$ of a given mode $k$ (defined as $a_k H_k \equiv k$, with $a_k \equiv a(t_k)$, $H_k \equiv H(t_k)$), and a later moment $t > t_k$. For tensor modes crossing during RD epoch, the resulting GW energy density spectrum retains the original shape

$$\Omega_{\text{GW}}^{(0)}(f) \bigg|_{\text{inf}} \approx G_k \Omega_{\text{rad}}^{(0)} \frac{H_{\text{inf}}}{m_p}^2 \left( \frac{H_{\text{inf}}}{H_{\text{max}}} \right)^2 \left( \frac{f}{f_{\text{RD}}} \right)^n \approx 10^{-16} \left( \frac{H_{\text{inf}}}{H_{\text{max}}} \right)^2 \left( \frac{f}{f_{\text{RD}}} \right)^n,$$  \hspace{1cm} (82)

with $n_t$ given by the original inflationary tilt.

If there is, however, a period of expansion after inflation for which the Universe energy budget is not dominated by relativistic species, the (quasi-)scale invariance of the original tensor spectrum is broken, as the inflationary GW energy density spectrum develops a significant tilt within the frequency range corresponding to the modes crossing the horizon during such period. We may detect or constrain this way the post-inflation expansion history (and hence the properties of the matter fields driving the expansion), by attempting to detect the relic SGWB of inflationary origin, see e.g. Refs. [898–905]. In order to see this, let us consider a cosmic epoch between the end of inflation and the onset of RD era, with effective EoS parameter $w \neq 1/3$.

In scenarios where the inflaton potential is $V(\phi) \propto \phi^3$, the inflaton oscillates around the minimum of $V(\phi)$, so that an effective EoS (averaged over inflaton oscillations) emerges following the end of inflation [111, 115], e.g. within the range $0 < w < 1/3$ for $p < 4$. For $p = 2$, it is well known that $w \approx 0$ emerges, as oscillations of a massive free field lead to the energy density redshifting on average (over one oscillation cycle) as $(\rho_\phi)_{\text{osc}} \propto 1/a^3$, analogous to that of nonrelativistic particle species. It is also possible that there is a stiff stage with $1/3 < w < 1$. The latter case can be actually achieved quite naturally if following the end of inflation the kinetic energy of inflaton dominates, either through inflaton oscillations [111] under a steep potential (e.g. $V(\phi) \propto \phi^p$ with $p > 4$), or simply by an abrupt drop of the inflationary potential.

Propagating the inflationary tensor modes through a nonstandard epoch starting immediately after the end of inflation leads to a GW energy density spectrum today, expressed as a function of present-day frequencies $f = k/(2\pi a_0)$, as [906]

$$\Omega_{\text{GW}}^{(0)}(f) = \Omega_{\text{GW}}^{(0)} \bigg|_{\text{inf}} \times W(f/f_{\text{RD}}) \times C_w \left( \frac{f}{f_{\text{RD}}} \right)^{2(3w-1)/(3w+1)} \approx \Omega_{\text{GW}}^{(0)} \bigg|_{\text{inf}} \times \begin{cases} 1, & f \ll f_{\text{RD}} \\ C_w \left( \frac{f}{f_{\text{RD}}} \right)^{2(3w-1)/(3w+1)}, & f \gg f_{\text{RD}} \end{cases},$$  \hspace{1cm} (83)

with $C_w$ just a numerical prefactor ranging $1 < C_w < 2^{5/2}/\pi \simeq 1.80$. The ratio -- for a fixed mode $k$ -- of the GW energy density spectrum to the EoS-$w$-background, scales therefore as $\rho_{\text{rad}}^{-1} (d\rho_{\text{GW}}/d \log k) \propto 1/a^4$. The growing function $\left( \Omega_{\text{GW}}^{(0)} \right)_{\text{inf}}$ -- as well as $\Omega_{\text{GW}}^{(0)}(f)$ -- for $w > 1/3$ (RD era) or $w < 1/3$ (infant universe), is well known to behave as proper GWs with an energy density scaling as $(d\rho_{\text{GW}}/d \log k) \propto 1/a^4$. The latter case can be achieved quite naturally if following the end of inflation the kinetic energy of inflaton dominates, either through inflaton oscillations [111] under a steep potential (e.g. $V(\phi) \propto \phi^p$ with $p > 4$), or simply by an abrupt drop of the inflationary potential.

Given the tilt in Eq. (83), scenarios with a stiff fluid dominated epoch between inflation and the onset of RD era are actually very appealing from an observational point of view, since they develop a blue tilt for tensor perturbations, $0 < n_t < 1$, at large frequencies [897, 899, 900, 902, 906–912], opening up the possibility of detection by upcoming direct detection GW experiments. The possibility of a stiff epoch is actually well-motivated theoretically in various early Universe scenarios, like Quintessential inflation [913–921], gravitational reheating [254, 255] (though with the caveats explained in Ref. [472]), Higgs-reheating scenario [561] and variants [922, 923]. The shape of the GW spectrum in these scenarios is controlled by $w$, $f_{\text{RD}}$, and $H_{\text{inf}}$, and the parameter space compatible with a detection by various experiments has been recently analyzed in Refs. [906, 912] (see also Ref. [22]). Consistency with upper bounds on stochastic GW backgrounds rules out, however, a significant fraction of the observable parameter space. This renders the signal unobservable by Advanced LIGO, independently of the parameter space $\{w, f_{\text{RD}}, H_{\text{inf}}\}$ [906]. The GW background remains detectable in experiments like LISA, but only in a small island of parameter space, corresponding
to contrived scenarios with a a low EoS $0.46 \lesssim w \lesssim 0.56$, and a high inflationary scale $H_{\text{inf}} \gtrsim 10^{13}$ GeV but low frequency $10^{-11}$ Hz $\lesssim f_{\text{RD}} \lesssim 3.6 \times 10^{-9}$ Hz (corresponding to a low reheating temperature $1$ MeV $\lesssim T_{\text{RD}} \lesssim 150$ MeV).

If there is an intermediate phase of MD (or, in general, a period with EoS ranging $0 \leq w < 1/3$), the transfer function of the inflationary GW spectrum in Eq. (83), develops instead a red-tilted high-frequency branch, as modes propagate through that phase. If the inflationary background could be observed by GW direct detection experiments, this could lead to determine the reheating temperature $T_{\text{reh}}$, since the end of the matter-like era would be determined by the start of RD era, and hence the pivot frequency $f_{\text{RD}}$ separating the two branches of the GW spectrum would inform us directly about the energy scale at the onset of RD epoch, see e.g. Refs. [903, 904, 924, 925]. Unfortunately, given the current upper bound on the scale of inflation, only futuristic experiments like BBO or DECIGO [926–928] may have a chance to measure the inflationary GW background today, as its amplitude is very suppressed, c.f. Eq. (82). Furthermore, unless upcoming CMB B-mode experiments make a detection of the inflationary background, the current upper bound on the scale of inflation is expected to be lowered by roughly an order of magnitude. Taking into account that, together with considering a natural reddening of the tilt at small scales (simply due to the progressive breaking of slow-roll as inflation evolves), it is very likely that the inflationary GW background will never be detected by direct detection GW experiments. Thus, the possibility to infer $T_{\text{reh}}$ with this technique seems rather difficult, unless B-mode experiments manage to detect the inflationary tensor modes within the next decade.

As discussed in Sec. 2.1, an EMDE following immediately after the end of inflation, can actually be considered a rather generic phenomenon, as long as the inflaton oscillates after inflation around its potential minimum, with the potential curvature dominated (at sufficiently small values) by a quadratic term. If this is the case, GWs produced by second order effects of the primordial curvature perturbation [929–932], can be enhanced at small scales, simply because of gravitational instability of scalar perturbations (as long as the EMD era lasts long enough, see e.g. Ref. [80] for recent numerical simulations on such scenarios). As a matter of fact, as discussed in Sec. 3.6, very large density perturbations at small scales may lead to the formation of PBHs, which potentially might explain the origin of (a fraction of) the DM. Early computations assuming a sudden transition between early MD and the onset of RD epoch [933], indicated that the induced GW background could be within the range of BBO/DECIGO, as long as the MD phase lasted until the reheating of the Universe at some low temperature $T_{\text{reh}} \sim 10^{6}$ GeV. Ref. [934] has, however, revisited the idea by incorporating the evolution of the gravitational potentials during a gradual transition from an early MD period to the RD era, as expected in more realistic models of a smooth transition with a timescale comparable to the Hubble time at those moments (as it is the case in perturbative reheating scenarios with constant decay rate). They conclude that the presence of an early MD era does not enhance the induced GW background as much as originally thought. For a sudden transition between MD and RD eras, however, an enhanced production of GWs occurs just after the reheating transition concludes, due to fast oscillations of scalar modes well inside the Hubble horizon as shown in Ref. [935]. They claim that this enhancement mechanism just after an early MD era is in fact much more efficient than any previously known enhancement mechanism during such MD period, leading to a detectable GW signal by BBO/DECIGO if the reheating temperature is in the range $T_{\text{reh}} < 7 \times 10^{-2}$ GeV or $20$ GeV $< T_{\text{reh}} < 2 \times 10^{7}$ GeV. In summary, the GW-induced effects due to an early MD period may strongly depend on how the reheating of the Universe took place. Also, more generally, the thermal history of the early Universe can be probed from the second order induced GW background [936, 937]. Going back to the case of a stiff epoch preceding the standard RD era, it is worth stressing that Ref. [938] has pointed out that in such a case the amplitude of the second order induced GW spectrum can be greatly enhanced, leading to a potentially observable signals at LISA and other planned observatories.

### 4.4.2. Cosmic strings

Cosmic strings are topological defects that can be produced in gauge theories undergoing phase transitions [939, 940] or fundamental strings stretched to cosmological sizes [941, 942]. While their evolution is in general very complicated, it does exhibit a scaling solution [943] in which the the total string density is a fixed small fraction of the total density. This behaviour is a consequence of the string network being chopped up into loops, which then oscillate and decay into GWs, though intercommutation. This behaviour was confirmed in numerical simulations [944–947] which were also crucial in fixing the loop number density which, in turn, is key in calculation of the resulting GW spectrum dominantly produced by loops.

For a recent review of the general formalism used to calculate the GW spectrum from cosmic strings we refer the reader to Ref. [948]. The key feature of these spectra is that a flat plateau develops, produced by the strings continuously emitting throughout the RD period. In the case of strings evolving in a nonstandard period of expansion [949–952] the spectra will change similarly to the case of an inflationary plateau discussed above. The modification will be visible above a characteristic frequency corresponding to the reheating temperature $T_{\text{reh}}$ [950]

$$f_{\text{reh}} = (8.67 \times 10^{-3} \text{ Hz}) \left( \frac{T_{\text{reh}}}{\text{GeV}} \right)^{1/2} \left( \frac{10^{-11}}{G \mu} \right) \left( \frac{g_{s}(T_{\text{reh}})}{g_{s}(T_{0})} \right)^{8/7} \left( \frac{g_{s,s}(T_{0})}{g_{s,s}(T_{\text{reh}})} \right)^{7/3} , \tag{84}$$

where $G \mu$ is the tension of the strings forming the network. If the modification is simply expansion driven before reheating by an energy constituent with a barotropic equation of state parameter $w$ the spectrum at higher frequencies will approximately follow Eq.(83), however, more subtle modifications such as a change in the number of degrees of
freedom will also leave a distinguishable feature in the spectrum at frequency \( (84) \) [950, 952]. The case of strings diluted by a short period of inflation is similar to the \( w = 0 \) case as the signal cannot go down more steeply than \( f^{-1} \) [953], however, strings produced close to the beginning of inflation can have a much more striking phenomenology as the stochastic background can be severely diluted leaving recently produced GW bursts as the main detection.

We show a few examples of possible standard and modified GW spectra in Fig. 16 together with expected sensitivities of planned GW observatories including LISA [955], ET [956, 957], AION/MAGIS [958–960], AEDGE [961] and SKA [962], as well as the existing PPTA [963] and LIGO [964] observatories.

### 4.4.3. Phase transitions

While study of production of GWs in a first order phase transition is a daunting task usually approached through lattice simulations [965–970] or through simplified analytical modelling [221, 971–976], typically the results are parametrised by just a few crucial parameters [429, 977, 978]. These are, firstly, strength of the phase transition,

\[
\alpha(T) = \frac{\Delta V(T) - T \frac{\partial \Delta V(T)}{\partial T}}{\rho R},
\]

which is the ratio of released vacuum energy to the radiation background. Secondly the time scale \( \beta \) which assuming exponential nucleation can be defined as

\[
\Gamma \propto e^{-\frac{S_3(T)}{T}} = e^{\beta(t_* - t)} \rightarrow \beta = -\frac{dS_3(T)}{dT} = H(T) T \frac{dS_3(T)}{dT}. \]

with the action of the nucleating bubble \( S_3(T) \) defined in Eq. (33). All the quantities marked with a * are evaluated at the transition temperature. There are three main sources of GWs in a phase transition: collisions of bubble walls [965, 968], sound waves produced by bubbles in the plasma [966, 967, 974] and turbulence ensuing after the transition [970–973]. All the expressions necessary to calculate the signal can be found for example in [222].

In the case of modified cosmology, the first obvious modification we have to make is to include any nonstandard components contributing to the expansion at the time of the transition in the abundance of GWs. If the extra components dominate the expansion during the transition this comes down to only a trivial modification of the standard formulas for \( \Omega_{GW} \) through

\[
\frac{\alpha}{\alpha + 1} = \frac{\rho V}{\rho V + \rho R} \rightarrow \rho V = \frac{\rho_{tot}}{\rho_{tot}} \alpha \left( \frac{H_{R*}}{H_*} \right)^2,
\]

where \( \rho_{tot} \) includes the new component dominating the expansion rate at the time of the transition and \( H_{R*} \) includes just the subdominant radiation component.

The last key parameter is the temperature at which the transition completes. This leads to another important modification with respect to the standard expressions which comes in through modified redshifting. GWs redshift in...
the same way as radiation: 

\[
\Omega_{GW,0} = \left(\frac{a_s}{a_0}\right)^4 \left(\frac{H_s}{H_0}\right)^2 \Omega_{GW,*} = 1.67 \times 10^{-7} h^{-2} \left(\frac{100}{g_*(T_{reh})}\right)^{\frac{1}{3}} \left(\frac{H_s}{H_{reh}}\right)^{\frac{3w+1}{3w+3}} \Omega_{GW,*},
\]

while the frequency scales as \( a^{-1} \):

\[
f_0 = \frac{a_s}{a_0} f_s = \frac{a_{reh} H_{reh}}{a_0} \frac{a_s H_s}{a_{reh} H_{reh}} \frac{f_s}{H_s} = 1.65 \times 10^{-5} \text{ Hz} \left(\frac{T_{reh}}{100 \text{ GeV}}\right) \left(\frac{g_*(T_{reh})}{100}\right)^{\frac{1}{3}} \left(\frac{f_s}{H_s}\right) \left(\frac{H_s}{H_{reh}}\right)^{\frac{3w+1}{3w+3}},
\]

where again the "reh" index denotes quantities at reheating. This prescription effectively means using the standard formulas at \( T_{reh} \) and including the extra redshift through the last bracket in equations above. There are several examples of interest:

1. For the case of extra energy constituent redshifting faster than radiation the reheating time and \( w \) are enough to calculate \((H_s/H_{reh})^2 = (H/H_{reh})^{2 \frac{3w+1}{3w+3}}\). Remembering that Eq. (87) enters the final GW abundance squared we see that the net result including redshift is still a suppression of the amplitude of the signal and increase of the frequency compared to the standard radiation dominated case. The upshot of this scenario is that the large Hubble rate at early times can help facilitate electroweak baryogenesis (see Sec. 3.2.1).

2. The second possibility is that the PT occurs in standard circumstances, however, its strong enough for the energy stored in the field itself to dominate the expansion as the field oscillates around its minimum effectively causing a short MD period before the field decays and reheats the Universe [222]. Thus, in this case Eq. (87) should not be used, yet the redshift is still modified according to Eq. (88) and Eq. (89).

3. The case of PTs in matter domination (or other constituent redshifting slower than radiation) is much less attractive as now both the redshifting and initial smaller abundance diminishes the signal. Thus the dominating MD period has to decay very soon after the PT to leave an observable abundance.

The upshot in the last two cases above is that reheating soon after the PT could also leave characteristic features in the spectrum [979] that would point specifically to such scenarios. We show an example of the resulting modification in Fig. 16 for the first two of the above cases.

5. SUMMARY

We have reviewed several causes and consequences of deviations from radiation domination in the early Universe – taking place either before or after Big Bang Nucleosynthesis – and the constraints on them, as they have been discussed in the literature during the recent years. In Sec. 2, we discussed the causes of nonstandard expansion and how they can broadly be grouped as modifications to the properties of the standard \( \Lambda \)CDM components, and the introduction of new components altogether. These two categories can be further refined into modifications to the background and/or perturbative evolution. As specific examples, we have discussed post-inflationary reheating, multiple stages of inflation, heavy particles, dark sectors, and moduli fields, as well as nonstandard post-BBN expansion history caused by e.g. decaying dark matter, nonstandard neutrino interactions, or early dark energy. In Sec. 3, we reviewed various consequences of nonstandard expansion eras including the effects such eras can have on models of dark matter, early Universe phase transitions and baryogenesis, models of inflation, and generation of the primordial curvature perturbation, as well as microhalos and primordial black holes. In Sec. 4, we summarized the known constraints on nonstandard expansion eras, as well as some future ways to put such scenarios into test. More specifically, we discussed constraints coming from BBN and the CMB, non-detection of microhalos and PBHs, and stochastic gravitational wave backgrounds.

The standard radiation domination that started at the end of inflation and lasted until the usual matter-radiation equality remains as a paradigm. However, several interesting proposals regarding various topics such as the generation of dark matter or matter-antimatter asymmetry during a nonstandard expansion phase have been recently made. We hope that this review on the possible causes and consequences of such eras and the known constraints on them helps in guiding further research on these topics.

\footnote{We approximate that \( g_*(T_{reh}) = g_{*,s}(T_{reh}) \).}
ACKNOWLEDGMENTS

R. Allahverdi is supported by NSF Grant No. PHY-1720174.
M. A. Amin is supported by the NASA ATP-theory grant NASA-ATP Grant No. 80NSSC20K0518.
A. Berlin is supported by the James Arthur Fellowship.
N. Bernal is partially supported by Spanish MINECO under Grant FPA2017-84543-P, Universidad Antonio Nariño grants 2018204, 2019101 and 2019248, and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreements 674896 and 690575.
C. Byrnes was supported by a Royal Society University Research Fellowship.
M. S. Delos is partially supported by NSF CAREER Grant No. PHY-1752752 and by a Dissertation Completion Fellowship from the University of North Carolina at Chapel Hill.
A. L. Erickcek is partially supported by NSF CAREER Grant No. PHY-1752752.
M. Escudero is supported by the European Research Council under the European Union’s Horizon 2020 program (ERC Grant Agreement No 648680 DARKHORIZONS).
D. G. Figueroa is supported by a Ramón y Cajal contract by Spanish Ministry MINECO, with Ref. RYC-2017-23493.
K. Freese is supported by the Jeff and Gail Kodosky Endowed Chair in Physics at the University of Texas, DOE grant DE-SC007859 at the University of Michigan, by the Vetenskapsrådet (Swedish Research Council) through contract No. 638-2013-8993, and the Oskar Klein Centre for Cosmoparticle Physics at Stockholm University.
T. Harada is supported by JSPS KAKENHI Grants No. JP19K03876 and JP19H01895.
D. Hooper is supported by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of High Energy Physics.
D. I. Kaiser is supported in part by the U.S. Department of Energy under Contract No. DE-SC0012567.
T. Karwal is supported by funds provided by the Center for Particle Cosmology at the University of Pennsylvania.
K. Kohri is supported by JSPS KAKENHI Grant No. JP17H01131 and MEXT Grant-in-Aid for Scientific Research on Innovative Areas JP15H05889, JP18H04594, JP19H05114, JP20H04750.
G. Krnjaic is supported by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of High Energy Physics.
M. Lewicki is supported by the UK STFC Grant ST/P000258/1 and by the Polish National Science Center grant 2018/31/D/ST2/02048.
K. Sinha is supported by DOE Grant DE-SC0009556.
T. L. Smith is supported by the Research Corporation, through a Cottrell Scholar Award, and NASA ATP grant number 80NSSC18K0728.
T. Takahashi is partially supported by JSPS KAKENHI Grant Number 17H01131, 19K03874 and MEXT KAKENHI Grant Number 15H05888, 19H05110.
T. Tenkanen is supported by the Simons Foundation.
J. Unwin is supported by NSF grant DMS-1440140 while in residence at the MSRI, Berkeley.
V. Vaskonen is supported by the UK STFC Grant ST/P000258/1 and by the Estonian Research Council grant PRG803.
S. Watson is supported in part by DOE grant DE-FG02-85ER40237 and NSF grant AST-1813834.
M. A. Amin, A. L. Erickcek, D. G. Figueroa, M. Lewicki, K. Sinha, T. Tenkanen and S. Watson acknowledge hospitality and support from KiTP-UCSB, where part of this work was carried out, supported in part by the National Science Foundation Grant No. NSF PHY-1748958.

REFERENCES

[1] S. Weinberg, *The First Three Minutes: A Modern View of the Origin of the Universe* (New York: Basic Books, 1977).
[2] A. A. Starobinsky, Phys. Lett. **B91**, 99 (1980).
[3] K. Sato, Mon. Not. Roy. Astron. Soc. **195**, 467 (1981).
[4] A. H. Guth, Phys. Rev. **D23**, 347 (1981), [Adv. Ser. Astrophys. Cosmol.3,139(1987)].
[5] A. D. Linde, *QUANTUM COSMOLOGY*, Phys. Lett. **108B**, 389 (1982), [Adv. Ser. Astrophys. Cosmol.3,149(1987)].
[6] A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
[7] A. D. Linde, Phys. Lett. **B129**, 177 (1983).
[8] D. Battefeld and P. Peter, Phys. Rept. **571**, 1 (2015), arXiv:1406.2790 [astro-ph.CO].
[9] R. Brandenberger and P. Peter, Found. Phys. **47**, 797 (2017), arXiv:1603.05834 [hep-th].
[10] G. Hinshaw et al. (WMAP), Astrophys. J. Suppl. **208**, 19 (2013), arXiv:1212.5226 [astro-ph.CO].
[11] C. Bennett et al. (WMAP), Astrophys. J. Suppl. **208**, 20 (2013), arXiv:1212.5225 [astro-ph.CO].
[15] R. Nguyen, J. van de Vis, E. I. Sfakianakis, J. T. Giblin, and D. I. Kaiser, Phys. Rev. Lett. 123, 171301 (2019), arXiv:1905.12562 [hep-ph].

[16] M. Bastero-Gil, M. Tristram, J. F. Macias-Perez, and D. Santos, Phys. Rev. D77, 023520 (2008), arXiv:0709.3510 [astro-ph].

[17] M. Bastero-Gil, J. Macias-Perez, and D. Santos, Phys. Rev. Lett. 105, 081301 (2010), arXiv:1005.4554 [astro-ph.CO].

[18] A. D. Linde, Nucl. Phys. B216, 421 (1983), [Erratum: Nucl. Phys. B223, 544(1983)].

[19] D. Bodeker and G. D. Moore, Phys. Rev. D53, 1784 (1996), arXiv:hep-th/9502041 [hep-th].

[20] D. Bodeker and G. D. Moore, Phys. Rev. D53, 1784 (1996), arXiv:hep-th/9502041 [hep-th].

[21] B. Carr, S. Clesse, and J. Garca-Bellido, (2019), arXiv:1904.02129 [astro-ph.CO].

[22] Y. Tada and S. Yokoyama, Phys. Rev. D100, 023537 (2019), arXiv:1904.02129 [astro-ph.CO].

[23] D. H. Lyth and E. D. Stewart, Phys. Rev. Lett. 75, 201 (1995), arXiv:hep-th/9405217 [hep-th].

[24] D. H. Lyth and E. D. Stewart, Phys. Rev. D53, 1784 (1996), arXiv:hep-th/9510204 [hep-th].

[25] R. Coleman, Phys. Rev. D15, 2929 (1977), [Erratum: Phys. Rev.D16,1248(1977)].

[26] C. G. Callan, Jr. and S. R. Coleman, Phys. Rev. D16, 1762 (1977).

[27] A. D. Linde, Phys. Lett. B100B, 37 (1981).

[28] A. D. Linde, Nucl. Phys. B216, 421 (1983), [Erratum: Nucl. Phys. B223,544(1983)].

[29] D. Bodeker and G. D. Moore, Phys. Rev. D53, 1784 (1996), arXiv:hep-th/9510204 [hep-th].

[30] B. Carr, S. Clesse, and J. Garca-Bellido, (2019), arXiv:1904.02129 [astro-ph.CO].

[31] J. Martin, T. Papapostolou, and V. Vassonkas, JCAP 1906, 024 (2020), arXiv:1907.04236 [astro-ph.CO].

[32] A. A. Starobinsky, JETP Lett. 42, 152 (1985), [Pisma Zh. Eksp. Teor. Fiz.42,124(1985)].

[33] J. Ellis, M. Lewicki, and J. M. No, JCAP 04, 003 (2019), arXiv:1809.08242 [hep-th].

[34] J. Ellis, M. Lewicki, J. M. No, and V. Vassonkas, JCAP 1906, 024 (2019), arXiv:1903.09642 [hep-ph].
