Early-Universe Model Building


1 Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
2 Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221, USA
3 Theoretical Physics Department, Fermilab, P.O. Box 500, Batavia, IL 60510, USA
4 Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA
5 William I. Fine Theoretical Physics Institute, School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA
6 Institute for Particle Physics Phenomenology, Durham University, Durham DH1 3LE, UK
7 Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA
8 Department of Physics, University of Toronto, ON M5S 1A7, Canada
9 Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87106, USA
10 High Energy Theory Group, Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA
11 IFAE and BIST, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
12 Université catholique de Louvain, Louvain-la-Neuve B-1348, Belgium
13 Department of Physics, University of California Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA and Santa Cruz Institute for Particle Physics, 1156 High St., Santa Cruz, CA 95064, USA
14 PRISMA+ Cluster of Excellence & Mainz Institute for Theoretical
Physics Johannes Gutenberg University, 55099 Mainz, Germany

15 School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK

16 Theoretical Physics Department, CERN, 1211 Geneva 23, Switzerland

17 Department of Physics & McGill Space Institute, McGill University, Montréal, QC H3A 2T8, Canada

18 Racah Institute of Physics, Hebrew University of Jerusalem, Jerusalem 91904, Israel

19 Maryland Center for Fundamental Physics, University of Maryland, College Park, MD 20742, USA

20 Department of Physics, Carleton University, Ottawa, ON K1S 5B6, Canada

21 Department of Physics and Astronomy, Rice University, Houston, TX 77005, USA

22 Leinweber Center for Theoretical Physics, Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

23 Berkeley Center for Theoretical Physics, University of California, Berkeley, CA 94720, USA

24 Theory Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

25 Department of Physics and Astronomy, University of Florence & INFN Florence, 50019 Sesto Fiorentino (FI), Italy

26 Laboratoire Univers & Particules de Montpellier (LUPM), CNRS & Université de Montpellier (UMR-5299), Place Eugène Bataillon, F-34095 Montpellier Cedex 05, France

27 Institute of High Energy Physics, Austrian Academy of Sciences, Georg-Coch-Platz 2, 1010 Vienna, Austria

28 Department of Physics & Astronomy, Wayne State University, Detroit, MI 48202, USA

29 Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany

30 TRIUMF, Theory Department, 4004 Wesbrook Mall, Vancouver, BC V6T 2A3, Canada

31 Department of Physics, Harvey Mudd College, Claremont, CA 91711, USA

32 Stockholm University and The Oskar Klein Centre for Cosmoparticle Physics, Alba Nova, 10691 Stockholm, Sweden.

33 Department of Physics, University of Notre Dame, IN 46556, USA

34 Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602, USA

35 FRIS, Tohoku University, Sendai, Miyagi 980-8578, Japan

36 Department of Physics, Tohoku University, Sendai, Miyagi 980-8578, Japan

37 DAMTP, Center for Mathematical Sciences, University of Cambridge, CB2 0WA, UK
Abstract

Theoretical investigations into the evolution of the early universe are an essential part of particle physics that allow us to identify viable extensions to the Standard Model as well as motivated parameter space that can be probed by various experiments and observations. In this white paper, we review particle physics models of the early universe. First, we outline various models that explain two essential ingredients of the early universe (dark matter and baryon asymmetry) and those that seek to address current observational anomalies. We then discuss dynamics of the early universe in models of neutrino masses, axions, and several solutions to the electroweak hierarchy problem. Finally, we review solutions to naturalness problems of the Standard Model that employ cosmological dynamics.
CONTENTS

1. Executive summary .................................................. 6

2. Dark matter models ................................................. 9
   2.1. Freeze-out DM .................................................. 10
       2.1.1. Classic freeze-out ..................................... 11
       2.1.2. Multiparticle dark sectors ......................... 12
       2.1.3. Kinematic effects ..................................... 13
       2.1.4. $3 \rightarrow 2$ and $4 \rightarrow 2$ interactions .......... 13
   2.2. Freeze-in DM .................................................. 13
   2.3. WIMPZILLA DM ................................................ 16
   2.4. Non-perturbative effects in dark sectors ................ 18
       2.4.1. Unstable bound states ............................... 18
       2.4.2. Stable bound states .................................. 20
       2.4.3. Intriguing dynamics during phase transitions .... 24
   2.5. Asymmetric DM ................................................. 26
   2.6. Atomic/Mirror DM ............................................. 29
   2.7. Sterile neutrino DM ......................................... 31
   2.8. Soliton DM .................................................... 35
   2.9. Axion(-like) DM .............................................. 39

3. Baryogenesis models .............................................. 43
   3.1. Leptogenesis .................................................. 44
   3.2. Electroweak baryogenesis .................................. 47
   3.3. WIMP baryogenesis .......................................... 52
   3.4. Baryogenesis by particle-antiparticle oscillations .... 55
   3.5. Mesogenesis .................................................. 57
   3.6. Axion baryogenesis ......................................... 60

4. Observational anomalies and cosmological model building .......................... 65
   4.1. $H_0$ and $\sigma_8$ tension .................................. 66
   4.2. 21 cm ....................................................... 73
1. EXECUTIVE SUMMARY

Fundamental physics has been incredibly successful in describing the early universe. For instance, the theories of electromagnetism and general relativity, supplemented with dark matter (DM), precisely explain the observed spectrum of the cosmic microwave background and the large scale structure of the universe. Furthermore, theories of the strong and weak interactions predict primordial abundances of light elements produced by Big Bang nucleosynthesis that are largely consistent with observations. However, these successful explanations are not yet complete from the particle physics point of view, as they are based on the so-called ΛCDM model, where the abundance of DM and baryon asymmetry are mere free parameters. It is not yet understood what the identity of DM is, how it was populated, and how the baryon asymmetry was created. Along these lines, one of the most important next steps in particle physics is to build models in order to explain the DM and baryon asymmetry, as well as to make predictions that can be confirmed or excluded by cosmological/astrophysical observations or terrestrial experiments. As we will review in part of this white paper, the ΛCDM model itself may need revisions in order to explain certain observational anomalies. This is likely to involve extensions to the Standard Model (SM) of particle physics and provide extra model-building opportunities.

Cosmology has rich implications for particle physics beyond just requiring the necessary ingredients discussed above. Indeed, the early universe may be considered as a “laboratory” equipped with a high-temperature bath. As such, its dynamics may allow us to probe particle physics models in a way that is impossible at the present time. For example, particles that are too heavy and/or too feebly interacting to be abundantly produced today can be easily produced at much earlier times, potentially leading to observable signatures (e.g., dark radiation in many extensions to the SM) or constraints on new particle physics models (e.g., gravitino problems in supersymmetric theories). Another important class of early-universe dynamics is the physics of phase transitions, which may create observable amounts of gravitational waves directly (first order phase transitions) or indirectly (production of topological defects.) An interplay between the early universe and fundamental particle physics also arises in theories where the dynamics of the early universe itself may be the key to solve some of the problems in the SM. Recently, several models have been proposed to explain the electroweak hierarchy problem through the evolution of scalar fields at early...
times. Ongoing and future pursuits along these lines will create new connections between particle physics and early universe cosmology.

Developments within observational cosmology play an increasingly important role in guiding model-building efforts. Unexpected data motivates us to find new creative explanations, thus adding to the model-building toolkit. Furthermore, more precise measurements of cosmological parameters continue to tighten and shape the viable landscape of new physics models, thus motivating new searches in terrestrial experiments. In the near future, upcoming observations of, e.g., the cosmic microwave background or 21 cm radiation may reveal hints of new physics that are otherwise inaccessible. Continued development of experimental technologies has also driven theoretical investigations into the nature of DM. For instance, low-threshold detectors sensitive to the sub-eV energy deposition of sub-MeV DM have fueled the exploration of cosmologically viable and detectable light DM candidates, while developments in precision sensors for ultralight bosonic DM across a wide range of mass scales have led to the identification of new cosmological targets within the larger parameter space.

Connections between the early universe and particle physics are becoming every increasingly tighter. In order to facilitate further development, in this white paper prepared for Snowmass 2021 we review particle-physics models of the early universe and their possible signatures. The remainder of this paper is organized as follows.

Secs. 2 and 3 discuss two firmly established cosmological problems: the nature and origin of DM and the baryon asymmetry of the universe. Here, we focus on bottom-up approaches and discuss freeze-out DM, freeze-in DM, WIMPZILLA DM, non-perturbative effects in dark sectors, asymmetric DM, atomic/mirror DM, sterile neutrino DM, soliton DM, axion DM, leptogenesis, electroweak baryogenesis, WIMP baryogenesis, baryogenesis by particle-antiparticle oscillations, mesogenesis, and axion baryogenesis. Models of DM and baryogenesis in solutions to the electroweak hierarchy problem are discussed in later sections. Sec. 4 discusses various observational anomalies that may require extensions to the SM: the $H_0$ and $\sigma_8$ tensions, the 21 cm observation by EDGES, the lithium problem, and unexplained features in small scale galactic structure.

Secs. 5-10 adopt a more top-down approach and review cosmological implications of various extensions to the SM that are motivated by other unsolved fundamental questions. Sec. 5 discusses models to explain the non-zero neutrino mass and mixing. Sec. 6 discusses axions
that may solve the strong CP problem and/or are predicted in string theory. Secs. 7, 8, and 9 discuss symmetry-based solutions to the electroweak hierarchy problems: supersymmetry, composite Higgs, and twin Higgs, respectively. Sec. 10 discusses the idea of dynamical vacuum selection, where the small electroweak scale and/or cosmological constant arises as a result of cosmological dynamics in the landscape of vacua. Abbreviations frequently used in this paper are summarized in Table I.

There is a plethora of white papers dedicated to early-universe physics in Snowmass 2021. Here we mention some that are closely related to early-universe model building. “Inflation: Theory and Observations” [1] reviews models of inflation, which are not discussed in this review, as well as their signatures in the CMB, large scale structure, and stochastic gravitational wave background. “New Ideas in Baryogenesis” [2] reviews models of baryogenesis, some of which are not covered in our paper. “Ultraheavy Particle Dark Matter” [3] outlines DM candidates with masses between 10 TeV and the Planck scale, with details on detection prospects. Secs. 2.3 and 2.4 of our paper also discuss models of heavy DM. “Primordial

2. DARK MATTER MODELS

The existence of DM is one of the most compelling pieces of evidence for physics beyond the SM. The most detailed understanding of DM arises from its gravitational imprints on cosmological observables, most notably from its effects on the acoustic peaks of the CMB angular anisotropies. Such observations have informed us of the abundance and general characteristics of DM when the universe was only $\mathcal{O}(10^5)$ years old. While its general particle physics description currently remains unknown, any DM candidate must be consistent with a plethora of terrestrial, astrophysical, and cosmological probes. At a minimum, any theory of DM must posit something that is non-relativistic, collisionless, feebly-coupled to normal matter, and stable. Even with these restrictions, the theory space is incredibly vast. Over the past decade the broadening of theory priors has enlarged the exploration of this theoretical landscape. This is due in part to: 1) the null results at large underground direct detection experiments and high-energy colliders and 2) the realization that the strong empirical evidence for DM motivates an examination of viable models beyond just those tied to other top-down motivations of new physics.

The sections below outline several classes of DM candidates with distinct cosmological origins. As DM is not the sole focus of this review, this list of models is by no means complete. Regardless, we hope that the discussion serves to illustrate the value of pursuing a multitude of model building directions to address DM’s cosmological origin. Secs. 2.1 and 2.2 discuss DM that was produced by the SM thermal plasma in the early universe, either through freeze-out or freeze-in, respectively. Sec. 2.3 introduces models of WIMPZILLAs,
FIG. 1. Thermal freeze-out processes that can set the abundance of DM in the early universe. The gray circle denotes a generic interactions while SM particles are shown in blue and dark sector particles are shown in black.

in which ultraheavy particles are created non-thermally, e.g., gravitationally due to the physics of inflation. Sec. 2.4 discusses a class of models in which DM arises as a result of non-perturbative dynamics, for example tied to the formation of bound states. Sec. 2.5 introduces scenarios where DM arises as a result of a primordial matter asymmetry, so-called asymmetric DM, while Sec. 2.6 discusses a natural consequence of such an asymmetry, where the DM is composed of dark atoms. Finally, Secs. 2.7, 2.8 and 2.9 discuss several other well-motivated models, where DM is composed of sterile neutrinos, solitons, or axions, respectively. DM models in solutions to the EW hierarchy problems are discussed in Secs. 7, 8 and 9.

2.1. Freeze-out DM

[Contributor(s): Jeff Dror]

Identifying the cosmological history of DM is one of the most profound goals for experimental physics in the 21st century. Early-universe probes track the evolution of DM through its gravitational imprint from redshifts of $O(10^3)$ until today. While these powerful observations indicate that DM must have formed prior to this epoch, we have yet to identify the production mechanisms and when it took place (as well as any potential non-standard
evolution of the universe from production until recombination). One compelling possibility
is DM underwent a thermal history analogous to particles in the visible sector; at early times
it contributed to the hot particle bath and, through its interactions with the known parti-
cles, its energy density was depleted until the two sectors “decoupled”. Such a history would
imply the abundance of DM is calculable from first principles, analogous to the abundances
of baryons and electrons (though their apparent particle-antiparticle asymmetry requires
adding a free parameter whose origin is yet to be identified (see Sec. 3).

For DM to thermalize with the hot particle bath requires a minimal assumption about
its interaction cross-section ($\sigma$) with the visible sector and the earliest temperatures of
the universe. In particular, the approximate condition for thermalization through $2 \rightarrow n$
scattering is for the interaction rate, $n \langle \sigma v \rangle$, to exceed the Hubble parameter before the
temperature ($T$) drops below its mass, $m_\chi$. In particular for a cross-section scaling as $\lambda^2/T^2$,
where $\lambda$ is a generic coupling strength, the condition is satisfied during radiation domination
if $\lambda \gtrsim \sqrt{m_\chi/M_{Pl}}$. Any sizable couplings satisfies this condition and will thermalize the
two sectors. At this point in the cosmic history, the co-moving number density would
be comparable to that of the visible sector bath, $\sim T^3$. Since we have not detected a
substantial DM energy density by Big Bang Nucleosynthesis, a thermal DM candidate must
have undergone some process to deplete its abundance prior to this epoch$^1$. In this section,
we summarize a selection of plausible depletion mechanisms and their consequences for
DM detection. We assume DM ($\chi$) thermalized in the early universe and its abundance is
subsequently depleted by collisions with other $\chi$-particles or antiparticles$^2$. The different
mechanisms are displayed in Fig. 1.

2.1.1. Classic freeze-out

In classic freeze-out (Fig. 1 A), the DM abundance is depleted through a two-to-two
scattering process until its rate drops below Hubble, at which time its co-moving num-
ber density is (again) conserved. To get an abundance in agreement with observations
requires a thermally-averaged annihilation cross-section, $\langle \sigma v \rangle \sim 10^{-26}$ cm$^3$/sec. For $\mathcal{O}(1)$

---

1 This argument breaks down if DM comes into thermal equilibrium with the SM after neutrino decoupling [9].
2 If $\chi$ annihilates by colliding with other dark sector states it can also exhibit rich phenomenology (see, e.g.,
Refs. [10–16]).
coupling sizes, this famously points to a DM mass around 0.1-1 TeV. As described, the classic freeze-out scenario is highly predictive, though it can be significantly modified through various tweaks to the framework above. The effort to summarize the “exceptions” to thermal freezeout was began with Griest and Seckel in 1991 [17] and has since grown into a body of literature, part of which makes up the rest of this subsection.

2.1.2. Multiparticle dark sectors

DM may not annihilate substantially with the visible sector; instead, it might annihilate or be converted into a secondary state ($\chi'$) which itself carries interactions with the SM. The evolution of DM depends sensitively on the mass of the secondary state. If $m_\chi \gg m_{\chi'}$, dark matter is in the secluded regime [18], where the annihilations proceed as in classic freeze-out requiring $\langle \sigma v \rangle \sim 10^{-26}$ cm$^3$/sec while the direct detection signal can be negligible (since it depends on the couplings of $\chi'$ to the SM instead of those of $\chi$). In contrast, the indirect detection rate is set by the $\chi$-annihilation rate leading to search prospects similar to those of classic freeze-out.

If $m_\chi \simeq m_{\chi'}$, thermal freeze-out of both $\chi$ and $\chi'$ must be studied in parallel and the dynamics depend on whether $\chi'$ annihilates or decays to deplete its abundance. If $\chi'$ annihilates, and the DM annihilation cross-section is similar to classic freeze-out (though to dark states) and the process is termed coannihilation [17]). Coannihilating DM can have direct detection and indirect detection rates comparable to that of classic freeze-out suppressed by a loop factor. If the secondary state decays, the dynamics further depend sensitively on the coupling of the $\chi'$ particle. If $\chi'$ remains in thermal contact with the SM during freeze-out, dark matter is again in the secluded regime, where the necessary annihilation cross-section needed to get the observed relic abundance is largely unchanged. If $m_\chi \simeq m_{\chi'}$ and $\chi'$ is long-lived (the co-decay limit [19]), the secondary state is likely to dominate the universe’s energy density prior to the decay. This in turn requires a correspondingly larger $\chi$ annihilation cross-section for the resulting dark matter density to match observations. This makes codecay a prime target for indirect detection searches. Semi-annihilation [20] is similar to secluded DM, but with a DM particle present in the final state of the $2 \rightarrow 2$ scattering process.
2.1.3. Kinematic effects

Forbidden DM [17, 21] is the same $2 \rightarrow 2$ process of classic freeze-out, but with the mass of the DM slightly below that of the outgoing SM state. This process, forbidden at zero temperature, has an exponentially suppressed rate proportional to $e^{-\Delta E/T}$, where $\Delta E$ is the energy threshold for the process. For the rate to be sufficient to deplete DM requires larger interaction sizes and hints at lighter DM. The exponential suppression typically leads to a negligible indirect detection signal but direct detection rates are not qualitatively unchanged.

2.1.4. $3 \rightarrow 2$ and $4 \rightarrow 2$ interactions

DM densities can be substantially influenced by $3 \rightarrow 2$ or $4 \rightarrow 2$ interactions, also known as a phase of cannibalism [22]. The dynamics during a cannibal phase depend on whether DM is chemically decoupled to the SM, kinetically coupled to the SM, or neither. If DM is both chemically and kinetically coupled, its bulk properties are unchanged over time. If it is only kinetically coupled, its number density drops, but its energy is transferred to the SM as the case for SIMPs [23], co-SIMPs [24], and ELDERs [25]. If its neither chemically nor kinetically coupled, then, while the number density is still dropping, the sector heats up as the mass energy is being transferred to kinetic energy. In this case, a thermal relic is too hot to form DM [22]. Cannibalization can also play the role of changing the evolution of the universe during annihilations [26, 27].

2.2. Freeze-in DM

[Contributor(s): Katelin Schutz, Saniya Heeba]

While freeze-out discussed in Sec. 2.1 is an attractive mechanism for DM production, it relies on assuming thermal equilibrium between the dark and visible sectors in the early universe. It is, of course, equally valid to consider the limiting case of extremely tiny DM-SM couplings such that the two sectors never equilibrate. In this case, the dark sector may be populated by the leakage of energy from the visible sector through sub-Hubble annihilations or decays of SM particles over time [28, 30]. This production mechanism is known as freeze-in because in many respects it is the opposite of freeze-out; the DM in non-thermal, many freeze-out processes can be time-reversed to provide a channel for freeze-in, and the
relationship between couplings and DM abundances have opposite behaviors for freeze-out and freeze-in. Due to the smallness of coupling strengths relevant for freeze-in, the DM candidate in such models is sometimes called a *Feebly Interacting Massive Particle* (FIMP). The freeze-in parameter space is an extreme limiting case because it represents the smallest relevant coupling the DM can have with the SM plasma that affects early-universe observables in any meaningful way. Historically, some classic examples of DM candidates with freeze-in production mechanisms include sterile neutrinos [31–34], singlet scalars [35], and various particles in supersymmetric models including sneutrinos [36–39], axinos and neutralinos [40–43], gravitinos [44–49], goldstinos [50], and photinos [51]. Freeze-in has since emerged as a more general mechanism that can be realized in a relatively model-independent way provided that the requisite couplings to produce the observed DM abundance are sufficiently small to avoid DM thermalization with the SM. The full theoretical exploration of the freeze-in mechanism for DM production remains an active topic that is ripe for further developments over the next decade.

One can estimate the upper bound on the effective freeze-in coupling by requiring the DM-SM interaction rate to always be sub-Hubble. This results in effective couplings of $\mathcal{O}(10^{-7})$ [52, 53] (although larger couplings may still be viable, for example, by considering cosmologies with small reheating temperatures [54]). Despite these small couplings, FIMP models display a rich phenomenology and can have complex early-universe dynamics. Such small couplings naturally dovetail with hidden sectors where there is a mediator that is feebly coupled to one or both of the DM and SM sectors. For example, freeze-in through the Higgs portal (i.e., mixing with a singlet scalar), the dark photon kinetic mixing portal, or other kinds of dark gauge bosons have all been extensively explored [30, 52, 55–60]. Aside from simpler FIMP models, recent studies have delved into freeze-in production of mediators which subsequently annihilate into DM particles (sequential freeze-in) [61], production through non-renormalisable operators (UV freeze-in) [62–66], freeze-in production followed by reannihilation or freeze-out in the dark sector [29, 52], and asymmetric freeze-in [67–69] which addresses the similarity between the cosmological DM and baryon densities. Additionally, since DM production by freeze-in is sensitive to a wide range of temperatures, a number of in-medium effects may play a significant role in DM production, including altered particle properties in a plasma, the spin statistics of relativistic quantum gases, and phase transitions [70–76]. State-of-the-art numerical codes for DM relic density calculations, such
as DarkSUSY [54] and micrOMEGAs [77], now have in-built freeze-in routines to account for some or all of these subtleties.

On the observational side, one might expect that FIMPs are intrinsically not detectable owing to their tiny couplings. This turns out to not be the case and freeze-in can be probed directly and indirectly using terrestrial experiments. The details of such detection schemes are quite dependent on the specific freeze-in model, but one quintessential example of this is the case of freeze-in via a light (sub-keV) kinetically mixed vector. This model has become a key benchmark for the sub-GeV direct detection program [78] because the same couplings determine the relic abundance and laboratory signals. In particular, the light mediator aids in the experimental observability of this candidate in direct detection experiments, since scattering will scale like $v^{-4}$ for velocity $v$, which is $v \sim 10^{-3}c$ at the Earth’s location in the Milky Way. Because of the IR-dominated temperature scaling of freeze-in production of DM via a light mediator, most of the DM is produced at late times, making this scenario relatively insensitive to initial conditions (unless the mediator is thermally populated at early times, which expands the parameter space [79] or unless the properties of the mediator change due to a phase transition [76]). Kinetically mixed vectors are the mediator of choice for this scenario because of strong stellar emission constraints on other low-mass mediators with the required couplings [56, 80]; in general, constraining the relevant mediators provides another indirect handle on freeze-in models and fixed-target and beam-dump experiments can help narrow down the viable parameter space [59, 81, 82]. Finally, one key signature of a feebly coupled dark sector is the appearance of particles with macroscopic lifetimes, which can be relevant to collider searches for displaced signatures in some FIMP models [83–85].

Cosmological and astrophysical avenues for detecting FIMPs are highly complementary to the terrestrial probes outlined above. Sub-MeV FIMPs are produced with a non-thermal phase distribution and may be constrained by studying their effects on cosmology. The portal responsible for creating FIMPs necessarily implies a drag force between the DM and the photon-baryon fluid before and during recombination, altering anisotropies seen in the CMB. Additionally, such light FIMPs inherit kinematics from heavier particles in the SM plasma and are thus quite boosted when they are born, which suppresses structure formation on small scales. Depending on the exact nature of the FIMP phase space, which is model-dependent, cosmological probes can be used to set a lower bound of $\sim 20$ keV on the FIMP mass [86–91]. Many FIMP models naturally lead to self-interactions that can be relevant for
the structure and abundance of DM halos at the present day \[73, 92, 94\]. Additionally, tiny FIMP couplings generically result in unstable DM candidates having long lifetimes. Such models can therefore be constrained (or act as targets for, in the case of various indirect detection excesses) indirect searches looking at decaying DM \[74, 81, 95–99\].

2.3. WIMPZILLA DM

[Contributor(s): Andrew Long]

It is often assumed that DM is a weak-scale particle whose relic abundance is controlled by thermal freeze-out (WIMP) as this scenario naturally embeds itself into weak-scale SUSY and other compelling theories of new physics at the TeV scale. Increasing the WIMP’s mass requires a commensurate increase in the WIMP’s couplings in order to maintain the desired relic abundance. Broadly speaking, WIMPZILLA is the idea that DM could be a stable elementary particle that is both super-heavy (with masses reaching up to the GUT scale) while also being very weakly coupled to ordinary matter (and possibly only coupled via gravity) \[100\]. Since WIMPZILLA DM cannot be produced by thermal freeze-out, other (non-thermal) production mechanisms are required \[101\]. Perhaps the most familiar example is inflationary gravitational particle production, and in fact gravitationally produced DM has almost become synonymous with WIMPZILLA. Other notable non-thermal production mechanisms include post-inflationary reheating or preheating, black hole evaporation, and cosmological phase transitions. In this short section, we offer a brief overview of WIMPZILLA DM: key elements of model building, phenomenological signatures, and experimental probes.

The presence of DM in our Universe, both today and 14 billion years ago during the epoch of radiation-matter equality, is inferred from cosmological and astrophysical observations that probe the DM’s gravitational interactions with ordinary matter and light. This abundance of data is consistent with the DM being a collection of massive elementary particles that interact with gravity in the same way that ordinary matter does (as opposed to a modified theory of gravity for DM). In many models where DM arises in a broader theoretical framework such as SUSY, the DM particles are predicted to have additional, non-gravitational interactions, which may play an important role for their creation in the early universe and detection in the laboratory. Nevertheless, for the time being, the data is consistent with DM that interacts only through gravity. The challenge for such models
is explaining the production of DM via gravity in the early universe, as well as providing observational channels for testing DM’s gravitational interactions in the laboratory.

The phenomenon of inflationary gravitational particle production (GPP) \[102, 103\], which occurs for quantum field theories in curved spacetime \[104, 105\], can explain how the DM was created in the early universe \[101, 106–108\]. Examples of GPP include Hawking radiation and the generation of the primordial density perturbations during inflation. The efficiency of GPP is related to the degree of non-adiabaticity experienced by the DM field. For example, consider a scalar field with mass \(m\) that is conformally coupled to gravity in a homogeneous Friedmann-Roberston-Walker cosmology with scale factor \(a(\eta)\) at conformal time \(\eta\). The Fourier modes with comoving wavenumber \(k = |\vec{k}|\) and comoving angular frequency \(\omega_k\) have the dispersion relation \(\omega_k^2 = k^2 + a(\eta)^2 m^2\). The non-adiabaticity in the evolution of these modes is estimated by \(\partial_\eta \omega_k / \omega_k^2 = (H/m)(1 + k^2/a^2 m^2)^{-3/2}\). For this model, a larger Hubble parameter \(H\) implies a larger non-adiabaticity, and consequently a larger abundance of gravitationally produced particles. GPP is most efficient either during inflation or at the end of inflation when the Hubble parameter is largest. Measurements of the CMB constrain the energy scale of inflation to be \(H_{\text{inf}} \lesssim 10^{14}\) GeV through the non-observation of primordial gravitational waves \[109\]. In typical models, inflationary GPP becomes less efficient for DM masses \(m \gtrsim H_{\text{inf}}\), implying that this phenomenon could explain the origin of DM for \(m \lesssim 10^{14}\) GeV.

Notable recent work has explored models of higher-spin DM \[110–115\], models of superheavy DM with \(m_{\chi} > H_{\text{inf}}\) \[116–119\], improved analytical techniques \[120–122\], and cosmological signatures such as isocurvature \[123–126\]. A complementary experimental effort is underway; the Windchime proposal \[127,129\] seeks to test the gravitational interactions of superheavy DM in the laboratory. The proposed experiment would entail a three-dimensional lattice of opto-mechanical sensors that could detect the passage of a compact object with a mass around the Planck scale via the mass’s coherent gravitational force.

Besides inflationary GPP, the production of superheavy DM may result from a direct coupling between the WIMPZILLA and the inflaton during the epoch of reheating, between the end of inflation and the beginning of radiation domination. WIMPZILLA DM may be produced directly as a decay product of the inflaton field or from the annihilation of inflaton decay products \[130,134\]. In such models, an accurate estimate of the relic abundance
requires a careful calculation of the primordial plasma’s thermalization and temperature evolution \cite{135,136}. Additionally, DM production may result from preheating \cite{137,139}, via non-linear dynamics of the inflaton and DM fields \cite{140}.

We close with a brief remark on WIMPZILLA production during cosmological phase transitions. These hypothetical events in our cosmic history are characterized by thermal symmetry breaking as the universe expands and cools. If DM couples to the symmetry-breaking field, then the phase transition can control the DM relic abundance. The case of a first order phase transition is particularly interesting, due to the rich non-linear dynamics of bubble nucleation and percolation. Bubbles can collide producing DM particles, they can act as filters to restrict the transmission of DM particles and limit their relic abundance, or they can modify DM’s properties like mass and couplings \cite{141,151}. Each of these scenarios has a rich phenomenology, particularly if the DM couples to the Higgs field during the EWPhT.

2.4. Non-perturbative effects in dark sectors

[Contributor(s): Pouya Asadi, Juri Smirnov]

The exact nature of the dark sector, in which the DM particle resides, is not known. However, in a broad class of scenarios, non-perturbative processes strongly affect the DM production process and its late-time phenomenology. Not only can there be different bound states in such sectors that can be viable DM candidates, but also other unstable bound states that can significantly affect the phenomenology. Other non-perturbative effects such as phase transitions can also give rise to intriguing dynamics. Such phenomena can occur both in confining dark sectors (and “strong” bound states) as well as dark sectors with only weakly-interacting long-range forces that can give rise to “weak” bound states. In what follows we review some recent progress in these threads of research.

2.4.1. Unstable bound states

Whenever the dark sector system is such that effectively long-range forces appear, non-perturbative phenomena are likely to be relevant in the non-relativistic limit. One well-known such effect is the so-called Sommerfeld enhancement \cite{152,153}. Here, at non-relativistic velocities, a potential between the incoming particles leads to the formation of off-shell bound states, and significantly affects the interaction amplitude. This can lead
FIG. 2. Electron-positron annihilation as analogy for DM annihilation. Line I) shows the tree level process, dominant at large velocities of incoming particles. Line II) shows the ladder diagrams of the Sommerfeld enhancement which is dominant at low energies. Finally, in line III) the capture photon emission is shown. This process dominates the annihilation process at low velocities and leads to interesting mono-chromatic signals for indirect detection searches in models of DM with EW interactions.

to a strongly enhanced late-time annihilation signal and needs to be taken into account for the freeze-out calculation.

On-shell bound-state formation is a related process that has been investigated more recently [154–156]. Its effect on the DM abundance [157, 158] and late-time signatures [159] become increasingly important with growing dark-sector coupling. A particularly interesting case is DM with EW charges. In particular for SU(2) representations larger than 3, the effective coupling strength is large enough to support effective bound state formation [159, 160].

Figure 2 shows an example for processes that affect the interaction rates of DM using the electron-positron annihilation analogy. The first line shows the tree-level annihilation process, which dominates the cross section at large energies of the incoming particles. The

<table>
<thead>
<tr>
<th>Process</th>
<th>Diagram</th>
<th>Cross-Section area</th>
</tr>
</thead>
<tbody>
<tr>
<td>I)</td>
<td>$e^+ e^- \rightarrow \gamma \gamma$</td>
<td>large velocity</td>
</tr>
<tr>
<td>II)</td>
<td>$e^+ e^- \rightarrow P^* \rightarrow \gamma \gamma$</td>
<td>small velocity</td>
</tr>
<tr>
<td>III)</td>
<td>$e^+ e^- \rightarrow P^* \rightarrow P \gamma$ $P \rightarrow \gamma \gamma$</td>
<td></td>
</tr>
</tbody>
</table>

![Diagram](image_url)
second line shows the diagrammatic representation of the Sommerfeld effect. Via the re-
summation of the ladder diagrams a non-relativistic potential between the particles can be
derived that, given an attractive interaction, enhances the annihilation probability. Finally,
the last line shows a related process, where an on-shell bound state forms, by emitting a
mediator and releasing the excess energy.

In the case that DM carries EW charges this mediator emission is particularly interest-
ing, since it provides a new mechanism for gamma-ray line from DM and is a new powerful
indirect detection search target \cite{157, 159}. Especially, in the case of ultra-heavy DM, the
photons from direct annihilation could be hindered in propagation, due to $e^+e^-$ pair cre-
ation \cite{161, 162}. However, the capture photon energies, which lie at the bound-state energy
scale, could still be detectable on cosmological distance scales \cite{159}. Note that at increas-
ing gamma ray energies even theoretical predictions become more involved \cite{163}, thus the
emission of capture photons at lower energies simplifies this theoretical problem as well,
providing a window on ultra-heavy DM candidates.

2.4.2. Stable bound states

Since almost all of the known mass of the universe is in the form of composite states,
either hadrons of the QCD (protons and neutrons) or weakly-bounded composite states such
as the Hydrogen atom, it is only natural to assume that the DM is also a bound state, either
formed by a confining hidden sector or another Abelian force, analogous to Hydrogen. This
is particularly further motivated by the fact that, even in the simplest form of such sectors,
there are numerous potential DM candidates. In what follows we review some of the recent
works on such models, point out interesting dynamics that arise in such sectors, as well as
novel signatures from such sectors.

- New dark forces, in particular if their interaction range is long, can lead to the forma-
tion of perturbative bound states. The recombination probability is large only at high
number densities in the early universe, which are typically only present in DM models
with a large matter-antimatter asymmetry \cite{164, 165}. For more details on such DM
models see, Sec. 2.6

- It is also possible that the DM is a stable bound state of a confining dark sector.
Depending on the details of the model, different symmetries can guarantee the stability of various DM candidates [166, 168]. Given their rich dynamics, the confining dark sectors can give rise to many interesting observable phenomena. Below we review different DM candidates in such sectors and some intriguing dynamics that can happen therein, see Refs. [167, 169, 170] for recent reviews of such models.

– Dark analogs of pions in SM can become stable kinematically or through various symmetries, giving rise to dark meson DM candidates, see for instance Refs. [166, 169, 171–181]. These dark mesons are composed of a dark quark and a dark anti-quark and are always bosons, regardless of the spin of dark quarks. The stability of such relics can be guaranteed in different ways. For instance, Ref. [169] used a flavor symmetry in the dark sector to achieve DM stability. Once gauged, the same symmetry can also naturally serve as a portal between the dark and the visible sector with various interesting signals [169]. This portal can further give rise to an ADM scenario (see Sec. 2.5) that simultaneously explains the observed abundance of SM and DM. Other manifestations of ADM can be built with dark meson DM models, see for instance Refs. [177, 180].

The stability of dark mesons can also be guaranteed via accidental symmetries, see for instance Ref. [166], or thanks to lack of non-gravitational interactions between the SM and the dark sector [169, 182]. In such models both dark mesons and dark baryons can be viable DM candidates.

Dark pion models can also provide a natural realization of SIMP [23] models thanks to the Wess-Zumino-Witten term [173]. Models with such freeze-out interactions have available parameter space only at masses around a GeV and lower. Depending on the ratio of dark quark mass and dark QCD scale, the mesons can have properties either similar to SM pions or heavier mesons such as charmoniums. There is also the possibility of the dark meson to be composed of a light and a heavy (compared to dark QCD scale) dark quark with interesting inelastic scattering signatures in direct detection experiments [171, 173].

– It is also possible that $N$ dark quarks charged under a dark confining $SU(N)$ form a color-neutral and stable baryon that can account for the observed DM abundance today, see for instance [166, 169, 174, 177, 182, 187]. Depending on $N$,
such baryons can be either fermions or bosons. Stability of such candidates can be guaranteed in many different ways, including by conservation of dark baryon number, accidental symmetries (e.g., see Ref. [166]) or lack of non-gravitational interactions with the SM.

In the latter scenarios the DM abundance is determined via freeze-in interactions. Despite the lack of non-gravitational interaction, the self-interaction among dark hadrons in such models can give rise to intriguing signatures, e.g., see Refs. [169, 182].

Similar to the case of dark meson models, it is possible to correlate the abundance of DM and SM in dark baryonic DM models as well, see for instance Refs. [177, 183].

It is also possible to charge constituent dark quarks under the SM gauge groups. Such models can give rise to interesting dynamics and signatures in a host of different experiments such as collider and indirect detection searches [183, 186]. Interesting signatures can exist in such models from decays of unstable hypermesons as well.

– Another potential DM candidate from confining dark sectors are glueballs, see for instance Refs. [188, 192]. The spectrum of glueballs in pure confining gauge groups has been studied extensively in the literature, see for instance Refs. [193, 194].

It is shown that many glueball states with different parity and charge-conjugation properties can exist in the spectrum of any confining gauge theory. Interactions of various glueball states within a pure Yang-Mills theory have been studied in Ref. [192]. It is shown that the internal interaction of glueballs can deplete most of the abundance of heavier glueball states into the lightest glueball state. If this lightest state is stable, it can be a viable DM candidate. Since no symmetries exists that can guarantee this particles stability, such a model will be viable only when there is no (or extremely weak) non-gravitational portal between the dark sector and SM.

Nonetheless, even if a non-gravitational portal that can make the lightest glueball unstable exists, heavier glueball states can still remain stable depending on the type of the portal [189, 190].
Similar to the dark baryons and dark meson models discussed above, the self-interaction between dark glueball states, which essentially happens with a cross section saturating the unitarity bound, can also give rise to interesting signals, see for instance Ref. [191]. It is also suggested that the dark glueballs can form macroscopic objects in the universe with interesting signals in gravitational lensing surveys [191].

Even if the glueballs are not stable, and thus not a viable DM candidate, they can still have substantial effect on the phenomenology of other dark hadron models. For instance, when they are long-lived enough they can inject substantial entropy into the SM bath after their decay, diluting the DM abundance, e.g. see Ref. [187].

Depending on the representation of dark quarks under the dark confining gauge group, other exotic hadrons can become viable DM candidates as well. An intriguing possibility is hadrons made of quarks in the adjoint representation and gluons of our QCD [178]. A follow up study applied this construction to a dark sector scenario [195]. A particularly interesting feature of this model is that DM undergoes a second stage of annihilation after the confinement phase transition, substantially affecting their relic abundance. This has also been used to derive a unitarity bound on extended DM masses [159]. (A more general study of such reannihilation epochs can be found in Refs. [196–198].)

Bound-state-induced effects and composite DM models can be viewed as UV-complete implementations of several mechanisms of early-universe scenarios reviewed in other sections of this paper. We anticipate with optimism, a number of experiments that will provide unprecedented sensitivity to the parameter space of DM models with bound states, within the next decade.

DM searches have extensively probed the DM mass range around the EW scale with astonishing accuracy. However, signals from dark sectors with non-perturbative dynamics naturally come with a much broader range of viable mass and interaction parameter values. Those models will be tested in (and provide well-motivated targets for) a multitude of upcoming experiments, both terrestrial and astrophysical searches [167 171 183 199–203], including novel collider searches [166 181 183 200 204 215].
So far the aforementioned experiments have not been able to detect any non-gravitational signals of DM. If this continues in the on-going and planned future experiments, devising new mechanisms that can open new parameter space will be strongly incentivized. Various novel dynamics in (confining) dark sectors can affect the relic abundance of DM and move the lamppost to hitherto unexplored mass ranges. In particular, with the recent surge of interests in gravitational waves, phase transitions in the early universe have been studied in more detail. Numerous studies have pointed out significant effects of (confinement) phase transitions on DM relic abundance and viable mass range, e.g. see Refs. [145–147, 149–151, 216–225].

In Refs. [151, 221] the flux tubes connecting isolated dark quarks of a confining dark sector were studied in detail and it was shown that during a first order phase transition fragmentation of these flux tubes produces many new boosted particles that mimic some aspects of QCD jets at colliders and significantly affect the asymptotic DM abundance. It has also been shown [149, 223] that in models with only heavy dark quarks (compared to the dark QCD scale), the confined phase bubbles nucleated during a first order phase transition can sweep the quarks into contracting pockets of the deconfined phase, giving rise to a second stage of annihilation that dramatically affects the final DM abundance. For high enough dark quark masses, and if no thermal contact to the SM sector was ever established, the pockets of the deconfined phase with quarks tapped inside become stable, giving rise to a viable macroscopic DM candidate dubbed “dark dwarfs” [224].

Similarly, in the presence of pre-existing non-zero dark baryon number, the first order confinement phase transition can also give rise to dark quark nuggets [216]. These macroscopic DM candidates can give rise to unique burst of electromagnetic radiation if they collide, as well as other signals in microlensing searches or in upcoming CMB experiments.

Non-confinement first order phase transitions have also been shown to have drastic effects on DM abundance. Consider a first order phase transition that takes a scalar field from zero vacuum expected value to a large one. If DM particles interact with this scalar and acquire their mass from its vacuum expected value, their varying mass near the higgsed phase bubbles acts as a potential barrier that prevents a non-negligible fraction of the DM particles from entering the higgsed phase bubbles [145, 147, 222]. Similar to the confinement scenario...
FIG. 3. A schematic depiction of the non-perturbative effect during a dark first order phase transition in Refs. [145–147, 149, 150, 222–224]. During the phase transition, SM particles and some of the dark sector particles can move into the low-temperature phase (blue bubbles), while other dark sector particles are trapped in the high-temperature phase (white background). As the phase transition proceeds (from the left to the right figure), these trapped particles are brought closer to each other in the high-temperature phase and go through a second stage of annihilation that depletes their abundance, and subsequently opens new parameter space for DM mass. In some parts of the parameter space the quarks trapped in the pocket can give rise to viable macroscopic DM candidates [147, 216, 224]. Figure is taken from Ref. [150].

discussed above, this will give rise to non-homogeneous over-densities of DM that can go through a second stage of annihilation, which substantially affects DM abundance. We schematically summarize these effect in Fig. 3. Such scenarios have been used for producing the observed baryon asymmetry of the universe as well [150].

Furthermore, in Ref. [219] it was shown that even if the scalar above rolls to a new vacuum, as oppose to tunneling during a first order phase transition, the change in DM particles mass can affect their abundance and viable mass range significantly.

The common feature of all these models is a new dynamic that suppresses DM abundance. Thanks to this suppression mechanism, the observed DM abundance today is predicted with lower interaction rates compared to vanilla freeze-out models. This lower interaction rate suppresses signals in various experiments and opens up new viable parameter space. Furthermore, as a result of the same suppression mechanism, parts of parameter space that was overclosing the universe in original freeze-out model becomes viable. Specifically, viable parameter space exists for masses beyond the often-quoted unitarity bound on thermal DM mass [226]. Existence of such viable parameter spaces strongly motivates an expansion of
the existing DM search programs to higher DM masses.

All in all, there are many reasons to believe further studies of non-perturbative dynamics in dark sectors can breed intriguing discoveries. Due to their rich dynamics, non-perturbative effects in dark sectors (in particular confining ones) can naturally give rise to manifestations of different DM models reviewed in other sections of this report, e.g. ADM models or strongly interacting DM models.

Studying these non-perturbative effects also has a natural synergy with many other research threads in high energy physics. This includes numerical works (such as lattice studies for better understanding of confining sectors dynamics and spectrum or better simulation of gravitational wave spectrum from a first order phase transition), as well as various phenomenological and experimental works (either in terrestrial or astrophysical searches). Furthermore, the study of BSM confining dark sectors beautifully complements various formal studies focusing on better understanding of the dynamics in confining theories.

2.5. Asymmetric DM

[Contributor(s): Hooman Davoudiasl]

The nature of DM remains unknown. Given the dearth of non-gravitational information on DM properties, it can be described by a multitude of models, covering a vast range of parameters. It is then often necessary to choose a guide for where to focus the efforts of theory and experiment, as the search for this mysterious substance continues. Perhaps one of the most intriguing potential hints regarding the identity of DM is encoded in its contribution to the cosmic energy density $\Omega_{\text{DM}}$, relative to that from the visible sector $\Omega_{\text{B}}$ dominated by baryons \[227\]:

$$R = \frac{\Omega_{\text{DM}}}{\Omega_{\text{B}}} \approx 5.$$  \quad (1)

As the properties of baryons and DM appear so different, one is then confronted with the question of why the above ratio is only modestly different from unity.

A possible way to answer the preceding question is to invoke a common origin for the baryon content of the Universe and DM (see Ref. \[228\] for an early suggestion in this direction). Since the former is set by an asymmetry, based on various direct and implicit empirical arguments, the DM abundance would also be taken to correspond to an asymmetry of DM over its antiparticle. We will refer to such a scenario, quite generally, as Asymmetric DM.
(ADM). A generic requirement of such models is the presence of sufficiently strong interactions that can lead to efficient elimination of the symmetric populations in both the visible and the dark sectors. For baryons, this is achieved by the presence of strong interactions in the low energy regime of QCD, while suitable new physics is required to do the same for ADM.

If similar asymmetric densities are achieved in both sectors, then the ratio in Eq. (1) implies that the mass of ADM is not far from the proton mass $\sim 1$ GeV. This is a typical requirement for ADM models (though variant models exist that do not lead to such a circumstance; see discussion below), but it is often assumed and not explained. In that sense, a more complete explanation of the above ratio $R$ would need to entail an explanation for why these masses are similar. We will come back to this point later. However, we point out that this mass scale is significantly different from the typical WIMP thermal relic scale $\sim 100$ GeV and affects the phenomenology in important ways. For one thing, it may be possible to produce the GeV scale DM using intense but low energy sources, like in fixed target experiments with electron or proton beams. Also, the low masses of ADM candidates make their signals challenging to uncover in nuclear recoil experiments that are often sensitive to WIMPs. We further note that indirect signals from annihilation processes in galaxies – typically accessible for thermal relic WIMPs – could be quite suppressed if the symmetric population is efficiently depleted in an ADM scenario (however, see Ref. [229], for intermediate scenarios where the symmetric component can be significant).

There is a multitude of models of ADM and performing a comprehensive survey of them requires a much more expansive article than this short contribution. However, there are a number of reviews that cover much of the basics and many key ideas; see for example Refs. [230] [232]. Also, there is no clear way to assign various ADM models to a few distinct classes. Nevertheless, there are some general features that broadly define different scenarios: (A) where a quantum number has an asymmetry that gets shared between the dark and visible sectors and (B) those that postulate a generalized “baryon number” that is preserved, with equal and opposite baryon and dark asymmetries.

Some of the earliest ADM models, based on EWSB from strong dynamics or “technicolor,” are of type (A), where $\sim$ TeV scale neutral and stable techni-baryons can be ADM [233] [234]. Here, thermal baryon number changing processes, or sphalerons, can distribute a shared baryon number between the dark and the visible sectors. Since techni-baryon changing
processes can stay in equilibrium below the mass $m_{TB}$ of the ADM, its number density can get suppressed $\propto e^{-m_{TB}/T}$, allowing for a weak scale ADM (see also Ref. [235]). This is a counter example to the typical GeV scale ADM. (See, e.g., Ref. [236], for recent work with ADM masses $\sim 60$ GeV and Ref. [223] with much larger masses $\sim 1-100$ PeV, using a different scenario. Alternatively, Ref. [237] considers very light scalar ADM.) However, strong dynamics is not necessary for this type of ADM scenario and it can be realized via a variety of other mechanisms [238, 239].

An early example of a baryon symmetric Universe of type (B), with and equal and opposite baryon number stored in a condensate, can be found in Refs. [240, 241]. In this model, ADM anti-baryons later emerge as scalars or non-topological bubbles. Another example is presented in Ref. [242]. Here, the dark and visible matter carry equal and opposite baryon number. In fact, ADM can scatter from ordinary baryons and destroy them, in this scenario. This scattering involves another ADM particle and a meson in the final state. Such a process leads to interesting detection possibilities at proton decay experiments, where the destruction of a nucleon by ADM mimics decay into a meson and a neutrino in standard decay scenarios [242, 244].

As mentioned earlier, in typical models the ADM mass scale is close to the proton mass, as asymmetries in the dark and visible sectors are taken to have a common origin and come out similar. Yet, in order to have a fuller explanation of the ratio in Eq. (1), one needs to explain the required $\sim$ GeV mass scale of the ADM. This can be achieved in models where the dark and visible sectors are connected by an exact or suitably broken $Z_2$ symmetry, causing the confinement scale of the “dark QCD” to be similar to that in the SM; see for example Refs. [245–248]. Similar scales of confinement may also be a result of unification of QCD and a dark confining sector at high scales [249].

There is a large variety of possible signals associated with ADM models proposed over the years, which may be accessed through cosmological, astrophysical, direct detection, and accelerator measurements. It is not feasible to survey all such phenomenology here. However, as mentioned earlier, ADM models often lead to DM populations that have suppressed annihilation rates in the late Universe, due to the scarcity of antiparticle states. This can lead to potential accumulation of ADM in astrophysical bodies, such as neutron stars, or the emergence of new exotic astrophysical objects made up of ADM states, like boson stars [250]. We will briefly discuss some of these features below. See also Sec. 2.6 for atomic/mirror DM,
which is a type of ADM with interesting astrophysical and cosmological properties.

Given some interaction with baryons, ADM can get gravitationally trapped in astrophysical bodies. Without significant annihilation, the ADM population grows and may collapse into a black hole. Formation of black holes inside, for example, a neutron star can lead to the disruption of the star. In the case of fermionic ADM, this process can be hindered by the Fermion degeneracy pressure of the dense ADM population \[251, 252\]. However, for bosonic ADM states this effect is not present and a Bose-Einstein condensate can form, leading to collapse into a black hole \[253\]. One can then use the observations of neutron stars to provide constraints on the interactions of the ADM with ordinary matter \[254\].

ADM can form new macroscopic astrophysical objects. Depending on the type of self-interactions of DM, such objects can have a range of masses, potentially reaching the solar mass scale \(M_\odot\) for fermionic ADM \[255\]. Here, Fermi degeneracy pressure provides a counterbalance to gravity. Such astrophysical objects can also be made of bosonic ADM. In the absence of Fermi pressure, the uncertainty principle prohibits the boson star from collapsing. With repulsive forces, masses of \(\sim 10^{11}M_\odot\) can be reached for these objects for ADM masses in the eV regime \[256\].

### 2.6. Atomic/Mirror DM

[Contributor(s): Francis-Yan Cyr-Racine]

Given that most of the visible matter in the Universe is in the form of neutral atoms, it is natural to ask whether all or part of DM could also exist in atom-like bound states. Such dark atoms have a long history in the literature (see Refs. \[257, 258\]) and occur naturally in mirror twin Higgs models (see e.g. Refs \[57, 212, 259–287\] and further discussion in Sec. \[9\]). In its simplest implementation, atomic DM \[164, 245, 288–308\] is made of two fermions of different masses oppositely charged under a dark \(U(1)_D\) gauge symmetry \[309, 310\]. More complex scenarios in which one or both constituents of the dark atoms are themselves composite particles (such as dark nucleons) are further discussed above in Sec. \[2.4.2\]. Similar to the visible sector, the presence of atomic DM generally requires a matter-antimatter asymmetry (see Sec. \[2.5\]) in the dark sector to set its relic abundance, although it is also possible for a symmetric component to survive \[311\], resulting in a mixture of darkly-charged DM \[312–314\] and dark atoms.

In the early universe at temperatures above the dark atom binding energy, the dark sector
forms an ionized plasma in which the particles charged under the dark \( U(1)_D \) are tightly coupled with the dark photon bath. Since this DR bath contributes to the overall radiation budget of the Universe \(^3_{15}\), the temperature of the dark sector needs to be somewhat lower than that of the SM to respect constraints on \( N_{\text{eff}} \) from the CMB and BBN. On the other hand, the presence of this extra radiation could potentially help alleviate the Hubble tension via the mechanism proposed in Ref. \(^3_{16}\). As long as the dark atoms are ionized, the large radiation pressure from the dark photons prohibits the growth of DM overdensities on scales smaller than their diffusion length. This large pressure support also leads to the propagation of acoustic waves within the dark plasma, akin to the more familiar baryon acoustic oscillations (BAOs) propagating in the visible primeval plasma. These so-called dark acoustic oscillations (DAOs) \(^2_{95}\) can leave important imprints on the cosmological distribution of matter on a variety of scales (see e.g. Refs. \(^3_{17-323}\)). Such signatures have been used to put constraints on these DM scenarios using an array of cosmological data \(^3_{24, 25}\).

The process of dark recombination, in which dark atoms are finally assembled into neutral bound states, plays an important role in the cosmological evolution of atomic DM. At that epoch, DM ceases to interact with the dark photon bath, hence allowing density fluctuations to start growing via gravitational in-fall. The efficiency of dark recombination depends sensitively on the value of the dark fine-structure constant \( \alpha_D \) and on the masses of the two particles involved \(^2_{274, 295, 326}\). Generally speaking, small values of \( \alpha_D \) or large masses result in inefficient dark recombination, leaving the dark sector largely ionized at all times, a scenario similar to that studied in Refs. \(^3_{12, 314}\). On the other hand, large values of \( \alpha_D \) or small masses generally result in extremely efficient dark recombination in which nearly all of the DM ends up in neutral atoms. How quickly the dark sector transitions from an ionized plasma to a mostly neutral dark gas plays an important role in determining the impact of DAOs on the matter power spectrum, with a very quick transition (compared to a Hubble time) resulting in large undamped oscillations in the the power spectrum, while a slow transition gives rise to a handful of severely damped oscillations \(^3_{18}\).

Much like standard baryonic gas, atomic DM can be collisionally excited, leading to dissipation in the DM sector. Such inelastic collisions can become particularly important in the later Universe when nonlinear structure formation leads to the formation of DM halos. Dissipation provides a mechanism for dark atoms to lose energy and momentum within
the potential wells of halos, leading to a modified halo structure and the possible formation of dark disks \cite{327,345}, though simulations are required to understand their predicted properties in detail. On even smaller scales, dissipation in the dark sector can lead to the formation of exotic objects such as dark stars \cite{346,349} and black holes in the mass gap of stellar evolution \cite{350,352}. Even in the absence of inelastic collisions, the typically large elastic cross section of dark atoms (due to their extended Bohr radius) means that they effectively form self-interacting DM \cite{296,353,355}. As such, atomic DM halos are expected to have cored density profiles if the self-interaction cross section is large enough for dark atoms to interact about once per Hubble time. For much larger cross sections, gravothermal collapse of atomic DM halos is also possible \cite{356,357}, leading to observational predictions that differ significantly from those of CDM.

If the dark photon kinematically mixes with the SM photon, a number of additional astrophysical and terrestrial signatures are generated. Atomic DM could lead to visible electromagnetic signals if it forms exotic compact objects like microhalos with enhanced annihilation \cite{358} or dark stars that capture SM matter from the interstellar medium \cite{347,348}. Atomic DM would accumulate in stars and contribute to white dwarf cooling, which is already providing strong constraints \cite{348}. Finally, atomic DM could leave signatures in direct detection experiments via the dark photon portal (see e.g. Ref. \cite{345,359}), subject to dark plasma effects from atomic DM capture in the earth \cite{345}.

2.7. Sterile neutrino DM

[Contributor(s): Marco Drewes, Bibhushan Shakya]

Neutrinos are the only particles in the SM that at least in principle possess the properties that one expects from DM (massive, neutral, and long-lived). However, within the standard scenario of three neutrino mass eigenstates and the observed values of the mixing parameters \cite{360,362}, explaining the total DM density in terms of thermal relic neutrinos would require the sum of neutrino masses \(m_i\) of \(\sum_i m_i \sim 11.5\) eV \cite{363} (clearly violating laboratory \cite{364} and cosmological \cite{365} constraints). Moreover, the neutrinos would be relativistic during cosmic structure formation, leading to clustering properties inconsistent with observation \cite{366}. Finally, neutrino DM would violate the “Tremaine-Gunn bound” \cite{367} on the phase space density of DM.

All of these problems could be overcome if there were an additional neutrino mass eigen-
state $N$ with a mass $M$ that is larger than a few keV and a mixing angle $\theta$ that is much smaller than the light neutrino mixing angles. Such an additional mass eigenstate (or several of them) can appear if there are any singlet ("sterile") fermions $\psi_s$ (with respect to SM quantum numbers) that mix with the SM neutrinos. Fermionic singlets can appear in many extensions of the SM because neutral fermions $\psi_s$ can generally mix with the SM neutrinos $\nu_L$ unless it is forbidden by some quantum number. The probably most prominent example are the right-handed neutrinos ($\psi_s = \nu^c_R$) but can also appear in other contexts (cf. e.g. [376]). It was originally assumed that $\nu_R$ should come with a Majorana mass $M$ far above the EW scale. However, since there are no large quantum corrections to this Majorana mass, in principle the allowed range of values for $M$ is very broad, leading to different phenomenological implications [377], and theoretical motivations for various values of $M$ have been brought forward [378], cf. Sec. 5.

The new mass eigenstate $N$ is an admixture of the SM neutrinos $\nu_\alpha$ (with $\alpha = e, \mu, \tau$) and $\psi_s$, $N \sim \psi_s + \sum_\alpha \theta_\alpha \nu_\alpha$, where the symbol " $\sim$ " indicates that the precise attribution of Lorentz- and spinor indices depends on the model (e.g. if one considers Dirac or Majorana particles). The mixing gives the almost sterile state $N$ a $\theta$-suppressed weak interaction and induces a non-unitarity of order $\theta^2$ in the light neutrino mixing matrix, which restricts the allowed values to $|\theta| \ll 1$ [379].

- **Indirect searches.** The $N$-particles are unstable, with the primary decay channel in the SM being $N \to \nu\nu\nu$ for small masses [380, 381]. In order to compose the DM, the $N$ must have a lifetime that exceeds the age of the universe, which imposes an upper bound $\theta^2_\alpha < 3.3 \times 10^{-4} \left( \frac{10^{3}\text{keV}}{M} \right)^5$ on $\theta_\alpha$ for given $M$. A stronger bound, however, comes from the non-observation of photons from the loop mediated process $N \to \nu\gamma$ [380, 381], which predicts a sharp photon emission line of energy $M/2$ from DM dense regions [382–384]. This practically restricts $M$ to the $\sim$ keV range unless one chooses all $\theta_\alpha$ so tiny that the mixing is practically negligible. The latter case is

---

3 The $\nu_R$ are often introduced in the context of the type-I seesaw mechanism for light neutrino mass generation [368,373]. The contribution from sterile neutrinos that comprise an $O(1)$ fraction of the observed DM density to the generation of neutrino masses is negligible [374], but they can be integrated in a type-I seesaw context with three or more flavours of $\nu_R$ [375].

4 Here, it is worth nothing that the Dirac mass between the left- and right-handed neutrinos is naively expected to be of the order of the Higgs vev, which results in $\theta$ being sufficiently large that $N$ is far too short-lived to be DM for any mass $M$. However, it is also technically natural for this Dirac mass to be vanishingly small, as needed for $N$ to be dark matter.
employed in many models, allowing for singlet fermion DM with much heavier masses.\footnote{These models are sometimes also referred to as “sterile neutrino DM”, though the connection to neutrino physics in this case would be only indirect, e.g. through some underlying neutrino mass generation mechanism.}

In 2014 an unidentified feature at 3.5 keV in the X-ray spectra of galaxy clusters \cite{385, 386} as well as Andromeda \cite{386} and the Milky Way galaxies \cite{387} was reported that can be interpreted as emission from the sterile DM decay $N \rightarrow \nu \gamma$. The DM interpretation of this signal has been the subject of an active discussion within the community ever since, see \cite{388} for a summary. A central topic of the ongoing discussion of the interpretation of data from XMM Newton \cite{389, 390} relates to the choice of background models \cite{391, 392} and statistical interpretation of the data \cite{389, 390}. Similar uncertainties exist in observations with NuStar \cite{395, 397}, for which the 3.5 keV lie at the edge of the sensitivity interval, making a quantitative estimate of the errors difficult. A central problem of all these studies lies in the limited spectral resolution of current X-ray telescopes, which is expected to improve considerably with the XRISM satellite \cite{398}.

All other bounds depend directly or indirectly on the mechanism through which the $N$ particles were produced in the early universe, and hence on the underlying model of particle physics. There are at least three different production mechanisms,

I) \textit{Thermal production through the $\theta$-suppressed weak interaction} \cite{31} tends to produce DM that is too warm to be consistent with observations unless it is resonantly enhanced due to the presence of lepton asymmetries that greatly exceed the observed baryon asymmetry in the primordial plasma \cite{399}. While the existing bounds on the lepton asymmetry \cite{400, 401} (which are considerably weaker than those on the baryon asymmetry \cite{402}) in principle allow for the generation of the required asymmetries \cite{403, 404}, in practice one needs a mechanism that generates them after the freeze-out of EW sphalerons at $T \sim 131$ GeV \cite{405}, which keep baryon and lepton numbers in equilibrium \cite{406}. One possible mechanism that can achieve this is the decay of heavier right-handed neutrinos \cite{407, 409} in the $\nu$MSM \cite{375, 410}. The resulting spectra in the latter case would be non-thermal \cite{403, 404} and may therefore in principle leave an observable imprint in the formation of structures.
II) Production from new (gauge) interactions is possible if the $\psi_s$ are charged under some extended gauge group beyond the SM. This is e.g. the case in left-right symmetric models [411-414] or in $U(1)$-extensions, where this mechanism has been studied in detail (cf. e.g. [415-419] and [420-422], respectively). New gauge interactions tend to bring the sterile neutrinos into thermal equilibrium, producing a too large amount of DM. The relic density production can be reduced to the observed value by a (drastic) change in the number of relativistic degrees of freedom $g_*$ in the plasma, entropy production in the decay of heavy particles, or self-annihilations after decoupling from the SM. The latter two possibilities have been studied in specific scenarios, e.g., in [415-417, 423, 424] and [425], respectively. If the decaying heavy particle has CP-violating interactions, it may in addition lead to successful baryogenesis [416].

III) The decay of heavy particles can generate sterile neutrinos with non-thermal (and sufficiently cold) momentum distribution. Various possibilities have been studied, including scalar singlets [32, 33, 426, 427] (c.f. e.g. also [64, 428-431]), charged scalars [432, 433], an additional Higgs doublet [433, 434], vector bosons [435, 436] or fermions [437]. The initial proposals [32, 33, 426, 427] assumed that the DM is produced during the freeze-out of the heavier particles. Production via freeze-in is also possible [429, 438-441], possibly leading to colder DM spectra [95], though a quantitative analysis may require a more detailed investigation of thermal corrections [71, 433] in this case.

IV) Gravitational production has been studied in the context of Einstein-Cartan gravity [442]. This mechanism is far less explored than the others and motivates further studies. The potentially strongest observational prospects come from

- **Phase space analysis**: Applying phase space considerations not only at present time but throughout cosmic history leads to a lower bound on $M$ [443-446]. For thermal production via the weak interaction the bound is $M > 2.79$ keV [443].

- **Structure formation**: With $\sim$ keV masses, the $N$ can have a non-negligible free-streaming length in the early universe, impacting the formation of structures in the

---

6 In ref. [447] a stronger (but disputed) claim has been made, arguing that phase space considerations and quantum statistics may favour a keV mass for the DM.
early universe on small scales. Observables in which this can be visible include the Lyman-\(\alpha\) forest, weak lensing constraints on the matter power spectrum, cosmic void properties, the number of high-redshift galaxies, counts of various objects on comparably small scales (galaxy halos, Milky Way satellites), 21 cm observations, \(N_{\text{eff}}\) and the shape of DM halos. The latter has attracted particular interest due to the possibility to address the so-called “core-cusp problem” [448]. Observational uncertainties come from the limited number of small scale objects that can be directly observed (since they are only observable in our nearby environment) and the lack of knowledge about the thermal history of the IGM (affecting the Lyman-\(\alpha\) forest). Theoretical uncertainties are dominated by the challenges that baryonic feedback poses to structure formation simulations. There is a vast body of literature addressing the impact of sterile neutrino DM on these observables, cf. [388, 449] for reviews and [450] for an update, which is a topic of active research. Observational prospects exist within specific scenarios for many of these observables and require further research.

Other constraints (e.g. from compact stars and supernovae) are either considerably weaker or suffer from considerable uncertainties, cf. [450]. In addition to these cosmological constraints, there are also a number of proposals to search for sterile neutrino DM in the laboratory [388, 449], with TRISTAN [451] currently under construction.

Overall, sterile neutrino DM remains a very active field of research, owing to the simplicity of minimal implementations, the potential connections to the mechanism of neutrino mass generation (cf. Sec. 5), and the possibility to find its imprints in astronomical and cosmological observations.

2.8. Soliton DM

[Contributor(s): Christopher Verhaaren]

The visible universe presents examples of simple elementary matter as well as composites of many particles combined. The cosmological DM may be of either type, or some combination of both. In this section we review solitonic DM. These scenarios are based on various soliton solutions of classical field equations. Such solutions correspond to collections of field quanta stabilized as a coherent unit of matter. These objects are interesting aspects of quantum field theory, but also provide a way to consider DM particles with masses at
or above the Planck scale but not the size of asteroids, planets, or stars. Thus, as definite realizations of macroscopic DM they can be used to explore the full range of possible DM masses. They also proved a concrete model of spatially large dark sector objects which can lead to novel DM search strategies [452].

We first consider extended field configurations that are stabilized by topological charge. These so-called topological defects can be produced nonthermally (and quite copiously) during phase transitions in the early universe and remain as the cosmological DM [453]. In some scenarios defects, like cosmic strings [454, 455], facilitate the production of other DM particles, but in this section we narrow our scope to the cases where at least some fraction of the DM is composed of the solitons themselves.

Our discussion begins with skyrmion DM. In any model where a symmetry $G$ is broken down to $H$, skyrmions can exist when the third homotopy group $\pi_3(G/H)$ of the symmetry breaking coset is nontrivial. The third homotopy group is significant because it pertains to mappings between three dimensional groups and the full three dimensions of physical space. In other words, the topological configuration of fields in the volume of space, rather than at the boundary, provides the topological stability.

Skyrmions may simply arise as a part of a dark sector. They can be linked to the SM by various portal interactions, such as the Higgs or neutrinos [456]. These connections can allow terrestrial experiments to search for these dark skyrmions [457].

In contrast, it may be that the skyrmion structure arises from within the SM. In [458, 459] it was shown that the EW sector of the SM, along with the usual skyrm term of the EW chiral Lagrangian can produce topological solitons. These solitons can make up the cosmological DM with masses in the few TeV range [460, 461].

In between these two cases is that of an extensions of the SM, specifically the Higgs sector, that can give rise to skyrmions. Models in which the Higgs is a pseudo-NGB are highly motivated by the well-known hierarchy problem related to the Higgs’ mass. Many of these models posit global symmetry breaking patterns with a nontrivial third homotopy group, such as $\pi_3(SU(N)/SO(N)) = Z_2$ for $N \geq 4$, which can allows the formation of TeV scale skyrmion DM [462, 463]. However, some care must be taken to determine the electric charge of the lightest skyrmion state [464]. The charge can depend on the UV completion of the model in many popular pNGB Higgs theories. Other construction, like $SO(N) \times SO(N)/SO(N)$ always lead to an electrically neutral lightest skyrmion state, which
could then be viable DM.

Magnetic monopoles are also topological defects, but in this case related to the second homotopy group $\pi_2(G/H)$ of a broken symmetry. The nontrivial topological configurations are not due to the orientation of fields in the bulk volume of space, but rather on the boundary at spatial infinity. Ever since Dirac showed that the existence of magnetic monopoles can explain the observed quantization electric charge they have been sought experimentally. However, this same characteristic predicts that any magnetic monopoles directly related to the electric charges of the SM must have a very strong coupling to SM photons, and clearly cannot make up the cosmological DM.

But these interesting topological objects could well be a part of the spectrum of a dark sector gauge theory. It has been shown that they can make up some fraction of the DM [465] along with other stable particles in the dark sector. Depending on the structure of the dark sector they may also be detectable through Higgs mixing [466]. A Higgs portal coupling can actually change the EW vacuum within the monopole core, leading to striking signatures in large volume detectors [467].

Kinetic mixing between our photon and the dark non-abelian gauge field [468] can also provide experimental access to dark monopoles. Abelian kinetic mixing between the SM photon and the dark photon that couples to the hidden monopole can provide the correct relic abundance for SIMP DM composed of dark monopoles [469]. It has been recently confirmed that heavy dark monopoles (masses greater than PeV) can also account for the entirety of the cosmological DM [470].

Intriguingly, dark monopoles, those uncharged under SM electromagnetism, can arise from Grand Unified Theories that do not posit a separate dark sector [471]. It has also been suggested that DM may be composed of the galaxy sized monopoles of a very weakly coupled dark $SU(2)$ gauge sector [472] [473]. In such a case a single monopole itself makes up the DM halo, rather than many individual monopoles together.

Connections between magnetic monopoles and axions have also been explored. This includes having an axion field connect the dark monopoles to the SM [474], providing novel ways to detect the monopole DM. The monopoles can also modify other cosmological aspects of a given model [475]. For instance, they can lead to a reduction in the relic abundance of QCD axions or isocurvature modes in cosmological density perturbations [476] [478] and prevent the persistence of domain walls [479].
If the dark $U(1)$ gauge symmetry under which the dark monopoles are charged is broken then they will be bound by strings of dark magnetic flux which affects the properties of a cosmological population. The flux string often makes them into unstable bound states, including interesting decaying DM candidates [480]. If there is a kinetic mixing between the SM photon and the dark photon then the dark magnetic monopoles obtain a magnetic charge under the visible photon that is suppressed by the mixing [481, 482]. This perturbative magnetic charge leads to novel phenomenology [483] including effects on magnetars [484] and galactic magnetic fields [485]. Such galactic constraints require the bound states be stabilized in some way, for example if there are two flavor of monopole so that the bound state constituents cannot annihilate. In these scenarios the monopole bound states also provide new avenues for terrestrial DM searches, including phase shifts in precise Aharonov-Bohm experiments [486].

So far we have considered solitons stabilized by topological charges. However, nontopological solitons have also been considered as DM candidates. A prime example are Q-balls [487], soliton solutions to certain scalar field theories. For a variety of potentials, the self-interactions of a complex scalar field can bind large numbers of quanta into a spherical ball. This configuration can have lower energy than an equal number of free scalars and is also stabilized by the conserved particle number $Q$.

These solitons have long been considered as DM candidates within SUSY theories [488–491], and are discussed in Sec. 7.4. However, SUSY is not essential to their DM quality. They have been investigated as self-interacting DM [492, 493] and also as a model for the DM halo itself [494]. Q-balls also allow for a variety of interesting connections to the SM [495], including solitons with EW symmetric cores [496], which enrich the possible DM phenomenology. They can even be combined with monopoles to produce DM with both topological and non-topological charges [497].

Like the topological solitons, nontopological solitons can be created during phase transitions in the early universe [226, 498], and their creation and subsequent evolution [499] has come to be called solitosynthesis. The production of these solitons is essential to understanding them as DM candidates [500], but can also have nontrivial implications for cosmic evolution. These include seeding further phase transitions [501, 502] and the production of gravitational waves [503–505].
2.9. Axion(-like) DM

[Contributor(s): Raymond Co]

Experimental evidence suggests that DM is cosmologically stable and feebly interacting with the SM. Ultralight axions naturally possess these necessary properties. In fact, axions are often predicted in the solutions to the unsolved puzzles of the SM. The most motivated examples include the PQ symmetry \[506, 507\] that solves the strong CP problem, the lepton symmetry \[508\] that provides the neutrino masses, and the flavor symmetry \[509\] that explains the fermion hierarchical masses and mixing. We refer to the pseudo-NGBs in spontaneous breaking of such global \(U(1)\) symmetries generically as the axions. More specifically, the QCD axion \[510, 511\] is defined when the axion interacts with the gluons as in the PQ mechanism, while other axions are called the ALPs. Oftentimes, the symmetry breaking scales are required to be large by the current experimental constraints and the axion masses may be naturally small due to the approximate \(U(1)\) symmetry, both of which automatically make the axions long-lived and feebly interacting. While axions are excellent DM candidates, a pressing question is the cosmological origin of the axion abundance which will be thoroughly addressed in this section.

The DM abundance can be conveniently expressed as the ratio of the energy density \(\rho_{\text{DM}}\) to the entropy density \(s\), where the observed abundance is \(\rho_{\text{DM}}/s \simeq 0.44\) eV. The required yield \(Y_{\text{DM}}\), defined as the number density over entropy density, \(Y_{\text{DM}} = n_{\text{DM}}/s = (\rho_{\text{DM}}/s)/m_{\text{DM}} = 0.44\) (1 eV/\(m_{\text{DW}}\)) is therefore much larger than unity for axions with mass \(m_a \ll 1\) eV. Since a full thermal abundance gives a yield at most \(\mathcal{O}(1)\), the production mechanism of ultralight axions must be non-thermal.

The non-thermal production of the axion is closely related to when and how the \(U(1)\) symmetry is spontaneously broken in the early universe. In other words, one needs to know the field evolution of the radial mode \(S\) of the complex field

\[
P = \frac{S}{\sqrt{2}} e^{iS},
\]

where the axion \(a\) resides in the angular direction and a non-zero value of \(S\) spontaneously breaks the symmetry. If such breaking occurs after inflation, the radial mode is initially trapped at the origin and later rolls towards the potential minimum at \(S = f_a\), with \(f_a\) the decay constant, spontaneously breaking the symmetry. At this time, the axion obtains random field values in different patches of the universe. If the symmetry is instead broken
before or during inflation, the field values of both modes will be homogenized by inflation up to possible quantum fluctuations. In what follows, we discuss how the axions can be produced in these scenarios.

In the post-inflationary scenario, there is an irreducible abundance from the misalignment mechanism \[512\] and the domain-wall network. Since the axion field is randomized, the axion field must have a spatial average of the misalignment angle \(\theta_i = a_i/f_a\) equal to \(\sqrt{\langle \theta_i^2 \rangle} = \pi/\sqrt{3}\). Due to the axion vacuum potential 

\[
V(a) = m_a^2 f_a^2 \left(1 - \cos \left(\frac{a}{f_a}\right)\right),
\]

this axion configuration leads to an abundance of the axion. The axion field will evolve according to the equation of motion

\[
\ddot{a} + 3H\dot{a} + V'(a) = 0,
\]

where the dots denote the time derivative and \(H\) is the Hubble expansion rate. The field value of the axion will be initially frozen until \(3H\) is comparable to the axion mass. To explain the observed DM abundance, this misalignment contribution alone predicts an axion mass of order \(m_a \simeq 30 \mu\text{eV}\) (or \(f_a \simeq 2 \times 10^{11}\ \text{GeV}\)) for the QCD axion and \(m_a \simeq 600 \mu\text{eV} (10^{12}\ \text{GeV}/f_a)^4\) for ALPs. In addition to axion misalignment, topological defects \[515, 516\], such as axion strings and domain walls, will also form. The decay of the axion-string and domain-wall network radiates axions, which also contribute to the DM abundance. An accurate determination of the total axion abundance in this scenario requires lattice simulations because of the complex dynamics involved from the \(U(1)\) symmetry breaking to the late-time oscillations. There have been extensive efforts dedicated to this numerical study \[517\] \[527\]. In the case where the domain wall number is larger than unity, the domain walls are stable and will overclose the universe. This issue is avoided if explicit \(U(1)\) breaking is introduced so that the domain walls decay and make an additional contribution to the axion abundance \[518, 528, 534\], which allows for a smaller \(f_a\) to still explain DM.

In the pre-inflationary scenario, the radial mode \(S\) plays a crucial role in determining the dynamics of the complex field \(P\) as well as the final abundance of the axion. The field value of \(S\) can be non-zero during inflation due to merely an initial condition, large quantum fluctuations, or dynamical relaxation by a mass larger than the Hubble scale during inflation.
In this case, the axion field is homogenized by inflation and the misalignment angle $\theta_i$ is a constant throughout the observable universe. The axion abundance, therefore, depends on the value of $\theta_i$. The observed DM abundance is reproduced by $m_a \simeq 10 \text{ } \mu\text{eV} \times \theta_i^{12/7}$ (or $f_a \simeq 6 \times 10^{11} \text{ GeV}/\theta_i^{12/7}$) for the QCD axion and $m_a \simeq 7 \text{ meV}/\theta_i^4 \times (10^{12} \text{ GeV}/f_a)^4$ for ALPs. Thus, a large decay constant, as motivated by string theory or grand unification of the gauge and $U(1)$ symmetries \cite{535, 536}, needs a small $\theta_i$ to avoid overabundance. Similarly, a small decay constant calls for an angle very close to the hilltop of the cosine potential, $\theta_i \to \pi$, in order to exploit the inharmonicity. In the simplistic scenario, a very small/large $\theta_i$ has to come from a tuned initial condition. However, a small angle may also result from the early relaxation of the axion field during or after inflation (a large angle is similarly achieved with a further phase shift of $\pi$ \cite{537-539}) if the axion mass is larger in the early universe \cite{540-543} or if inflation lasts a very long time \cite{544-547}.

In the above discussions of post- and pre-inflationary scenarios, we have assumed a radiation-dominated universe after inflation so that no entropy is produced during or after the axion production. If there exists an early matter-dominated era, e.g., due to the inflaton or a moduli field such as the radial mode $S$ itself \cite{42, 548-551}, the universe undergoes a reheating period, which generates entropy and dilutes the DM abundance. This opens up parameter space with a larger $f_a$. On the other hand, if the axion is produced during an era of enhanced Hubble without subsequent dilution, such as a kination era, the axion abundance is enhanced instead because the onset of axion oscillations and accordingly the energy redshift are delayed \cite{552}.

Interesting and non-trivial dynamics also occurs if the radial mode $S$ takes on an initial field value much larger than $f_a$. In the limit of a $U(1)$-symmetric potential, $S$ starts to oscillate about the origin of the PQ potential when $3H$ is comparable to the mass of $S$. The oscillation of $S$ leads to an oscillatory mass of the mode orthogonal to the oscillation direction via a self-interaction term in the potential of $P$, e.g., a quartic term of $P$. (Such an interaction term must be present in order for the $U(1)$ to be spontaneously broken.) This oscillatory effective mass leads to efficient production of fluctuations by so-called parametric resonance \cite{137, 138}. As a result, the axion fluctuations are generated and can be cold/warm DM \cite{553, 554} when the momentum is sufficiently redshifted. In fact, parametric resonance can also occur when $S$ starts to oscillate from the origin towards the minimum after being thermally trapped \cite{555}.
A perfectly $U(1)$-symmetric potential may not be realistic to assume. For instance, quantum gravity is conjectured to not respect global $U(1)$ symmetries $[556-560]$, in which case explicit breaking is expected in the form of higher dimension operators $[561-564]$. In the case of the QCD axion, whose PQ symmetry is anyway explicitly broken by the QCD effects, the PQ symmetry is at best understood as an accidental symmetry, and thus explicit breaking from other sources is plausible. The quality of the axion solution to the strong CP problem requires such explicit PQ breaking terms to be operators of a sufficiently high dimension. Although these higher dimension operators can be negligible when $S = f_a$, they may play a significant role when $S$ has a large value in the early universe, leading to unsuppressed explicit breaking. Since this gives the axion a large initial mass around the onset of the motion of $P$, the trajectory is no longer a radial oscillation alone but also involves the angular rotation. This initiation of the rotation is analogous to that proposed in the Affleck-Dine mechanism $[565]$. The resultant large velocity in the axion direction can lead to an enhanced axion abundance via the kinetic misalignment mechanism $[554, 566]$. The enhancement arises because the axion energy is dominated by the kinetic energy rather than the potential energy at the temperature the axion would start to oscillate in the conventional misalignment mechanism. Kinetic misalignment reproduces the observed DM abundance with a smaller decay constant. These dynamics of axion rotations is also intimately connected to a baryogenesis mechanism called axiogenesis $[567]$ to be discussed in Sec. 3.6, which along with kinetic misalignment would allow a correlated prediction of the axion mass and decay constant.

Other types of explicit PQ breaking that are important at high temperatures may also alter the axion evolution significantly. For example, the axion mass may be enhanced due to the Witten effects $[568]$ involving hidden monopoles $[568, 569]$. In this case, the axion may go through early oscillations to relax to a different minimum from the vacuum one, which generically does not lead to a prediction of the DM abundance $[476, 478, 570, 571]$. However, the abundance may be suppressed if the axion adiabatically tracks the temperature-dependent minimum of the potential $[476, 478, 571]$. If additional explicit breaking is aligned with that from the QCD, such breaking can remain effective in the vacuum without spoiling the solution to the strong CP problem. In particular, Ref. $[572]$ shows that the axion mass can be suppressed if there are $N$ copies of the SM with a $Z_N$ symmetry. The abundance can be enhanced in this case because the peculiar temperature dependence of the axion potential.
can lead to early oscillations that trap the axion in an early misaligned minimum, which delays the onset of oscillations around the vacuum minimum [573]. It is worth noting that \(N > \mathcal{O}(10)\) is necessary to have a significant effect on the parameter space. Refs. [574, 575] pointed out that, even if additional explicit breaking is misaligned from that of the QCD and is as small as allowed by the bound on the neutron EDM, the axion may still be trapped to a wrong minimum at high temperatures until the potential from QCD effects dominates. This scenario can enhance or suppress the axion abundance depending on the initial condition.

Lastly, extra fields in models that couple to the axion can also alter the axion abundance from the misalignment contribution. If the axion couples to a dark photon with a field strength \(F_D\) and a dark fine structure constant \(\alpha_D\) via \(\mathcal{L} \supset c_D(\alpha_D a/8\pi f_a)F_D\tilde{F}_D\), production of dark photons via tachyonic instability may occur for \(c_D\alpha_D \gg 1\). This tends to suppress the axion abundance [576] due to its transfer of energy to the dark photons, which will behave as a negligible amount of DR. However, the suppression factor is limited due to the backreaction on the axion according to the lattice simulations [577]. The large coupling \(c_D\alpha_D > \mathcal{O}(30)\) also requires additional model building. Ref. [578] studied a different model where the axion is simultaneously coupled to the dark photon and the SM photon with a field strength \(F\) via \(\mathcal{L} \supset (a/2f_a)F_D\tilde{F}\). If a background magnetic field is present in this model, Ref. [578] finds an enhanced axion abundance due to a delayed onset of oscillations.

### 3. BARYOGENESIS MODELS

In the hot expanding universe, if it were not for the primordial excess of baryons over anti-baryons, there would not be as much of matter as we observe today. The size of the baryon asymmetry of the Universe is precisely determined by BBN and the CMB, and the up-to-date estimate is [579]

\[
Y_B \equiv \frac{n_B}{s} = 8.7 \times 10^{-11},
\]

where \(n_B\) is the baryon number density and \(s\) is the entropy density of the Universe. The origin of this baryon asymmetry is one of the most important questions in particle physics and cosmology. Mechanisms to generate baryon asymmetry are called baryogenesis.

Three conditions must be satisfied for successful baryogenesis [580]: 1) violation of baryon number, 2) violation of C and CP symmetry, and 3) departure from thermal equilibrium. In
many models of baryogenesis, the required violation of CP symmetry and baryon number have experimental implications. Departure from thermal equilibrium may imprint cosmological signals such as gravitational waves.

In this section, we review several representative models of baryogenesis: leptogenesis, EWBG, WIMP baryogenesis, baryogenesis by particle-antiparticle oscillations, mesogenesis, and axion bagyogenesis. Baryogenesis in solutions to the EW hierarchy problem is discussed in Secs. 7 and 8.

3.1. Leptogenesis

[Contributor(s): Brian Shuve and Jessica Turner]

Neutrino mass-scale data \cite{581, 582} indicates that neutrino masses are significantly smaller than the masses of other SM fermions. One possible explanation is the type-I seesaw mechanism \cite{368, 370, 371, 583} which augments the SM by at least two right-handed neutrinos (RHNs) with masses $M_N^i$:

$$L = i N_i \phi N_i - \bar{\nu}_\alpha Y_{\alpha i} N_i \Phi - \frac{1}{2} N_i^C M_N N_i + \text{h.c.},$$

(6)

where $Y_{\alpha i}$ is the Yukawa matrix, $N_i$, $\nu_\alpha$, and $\Phi$ denote the RHNs of generation $i$, SU(2)$_L$ leptonic doublets of flavor $\alpha$, and Higgs doublets, respectively, with the negative hypercharge Higgs doublet defined as $\Phi = i \sigma_2 \Phi^*$. After EWSB, a Majorana mass term for SM neutrinos is induced, $m_\nu \approx Y M^{-1} Y^T v^2$, where $v$ is the VEV of the Higgs. In addition to providing a simple explanation of neutrino masses, the type-I seesaw satisfies Sakharov’s conditions \cite{580}: lepton number violation is provided by $M_N$, the complexity of the Yukawa matrix can mediate CP-violating interactions and the RHNs can decay out of equilibrium in the early universe. As the RHNs are singlets under SM gauge interactions, their masses are not constrained by the EW scale and can be large. Taking $m_\nu \sim 0.1$ eV and $Y v \sim 100$ GeV, a lepton number violating scale of $M_N \sim 10^{14}$ GeV is predicted. In the early universe, a population of these heavy RHNs can be produced through scatterings processes involving SM particles at $T \gtrsim M_N$. Their $CP$-violating and out-of-equilibrium decays can generate an asymmetry in SM leptons via $N_i \rightarrow L_\alpha \Phi^*, \bar{L}_\alpha \bar{\Phi}$, which in turn is partially converted into a baryon asymmetry via EW sphaleron processes \cite{406}. This mechanism is known as thermal leptogenesis \cite{584}.

If the RHN mass spectrum is hierarchical ($M_{N_1} \ll M_{N_2} \lesssim M_{N_3}$), then the majority of the
baryon asymmetry is produced by the decays of the lightest RHN, $N_1$. Without tuning of the Yukawa matrix, $M_{N_1}$ may be as low as $10^9$ GeV for such a hierarchical spectrum [585]. In the case that $M_{N_1} \gtrsim 10^{12}$ GeV, then the leptons and anti-leptons that couple to the lightest RHN maintain their coherence as flavor superpositions throughout the leptogenesis era, $T \sim M_{N_1}$. However, if leptogenesis occurs at lower temperatures, then scatterings induced by the SM charged lepton Yukawa couplings occur sufficiently fast to distinguish the different lepton flavors and decohere into their flavor components. The dynamics of leptogenesis must then be described in terms of the flavor states [586, 589]. While most thermal leptogenesis studies focus on the scenario where decays of $N_1$ dominantly produce the lepton asymmetry, an $N_2$-dominated leptogenesis scenario is possible [590] if there is a third RHN in the spectrum and the $N_1$ washout is not large. The scale of thermal leptogenesis can be lowered to $M_N \sim 10^6$ GeV, if flavor effects and tuning of the Yukawa matrix are taken into account [591]. However, a significant lowering to TeV-scale leptogenesis occurs if the mass differences between the heavy neutrinos are comparable to their decay widths, and this scenario is referred to as resonant leptogenesis [592]. If the RHN masses are $M_N \sim 1000$ TeV with a few TeV-scale mass splitting, RHN radiative corrections to the Higgs potential can account for the origin of the EW scale and provide successful resonant leptogenesis [593, 594].

In the above scenarios, leptogenesis occurs via the freeze-out (or departure from equilibrium) of RHNs in their decays. Alternatively, leptogenesis can occur via the freeze-in (or production) of out-of-equilibrium RHNs. This mechanism is known as Akhmedov-Rubakov-Smirnov (ARS) leptogenesis [410, 595]; see Ref. [596] for a recent review, as well as references therein. ARS leptogenesis allows for the effective generation of a baryon asymmetry at high temperatures, even for RHN masses at the GeV scale or below. In this mechanism, the production, coherent propagation, and subsequent scattering of RHNs generate asymmetries in individual SM lepton flavors. While the total lepton asymmetry is zero at leading order, flavor-dependent washout effects can convert the flavor asymmetries into total lepton asymmetries, which sources a baryon asymmetry via sphalerons. Interestingly, the generated asymmetry typically originates from lepton-number-conserving interactions, although lepton-number-violating terms have important implications for the asymmetry with highly degenerate and heavier ($\gtrsim$ GeV) RHNs [597, 598]. ARS leptogenesis is a central component of the $\nu$MSM [375, 410], a popular model in which the problems of DM, neutrino masses, and
baryogenesis are all solved with only three new, sub-weak-scale states (two responsible for leptogenesis and one DM). As a model of freeze-in leptogenesis, the mechanism is sensitive to the presence of other, non-minimal interactions of the RHNs, which can substantially suppress the asymmetry [436, 599].

There is a significant overlap between the parameter space of ARS leptogenesis and resonant leptogenesis; indeed, freeze-in leptogenesis can be effective for RHNs at or above the weak scale [600, 601], and low-scale RHNs often give both freeze-out and freeze-in contributions to the asymmetry [597, 598]. A recent study has provided a unified treatment of both phenomena [602]. Studies of ARS leptogenesis with three RHNs involved in asymmetry generation have shown that the parameter space expands relative to the minimal scenario [603–605], and the RHNs no longer need to be degenerate in mass.

From an EFT perspective, the existence of neutrino masses originates from the Weinberg operator, $O_{\alpha\beta} = (L_\alpha\tilde{\Phi})(L_\beta\tilde{\Phi})/\Lambda$ in Weyl spinor notation. While the Type-I seesaw mechanism provides a UV completion of this operator, there are many UV completions. If we restrict ourselves to tree-level UV completions, then there remain two possible scenarios: models with a new SU(2) triplet scalar (Type-II) [373, 606–608], or fermion (Type-III) [609]. As in the Type-I seesaw, the interference of tree and loop diagrams leads to an asymmetry from the decays of these new states [610, 611]. As with resonant leptogenesis, the new states can be near the EW scale for quasi-degenerate spectra. The phenomenology of these models is dramatically different from the Type-I seesaw due to the EW charges of the new states. For a comprehensive review, see Ref. [612].

Leptogenesis in Type-I seesaw models can be tested when the RHNs have masses $\lesssim$ TeV [613]. In the minimal model, RHNs are produced through their mixing with SM neutrinos, leading to production in charged-current weak interactions at colliders and beam-dump experiments [614–620]. The smallness of the RHN coupling at low masses typically leads to a long lifetime for these particles and displaced vertex searches can be promising modes of discovery. Nevertheless, testing the entire seesaw parameter space is challenging due to the tiny couplings involved, particularly for RHN masses near or above the weak scale.

Extensions of the minimal model, for instance with an extra U(1) gauge group coupled to $B-L$ or additional scalars, allow for larger production rates of RHNs than predicted by the naïve seesaw. Such models also can dynamically explain the mass degeneracy and/or coupling alignments found in resonant and ARS leptogenesis [621]. In such models, there
can be a significant rate of RHN single or pair production in \(Z', W', \) SM Higgs, meson, or other scalar decays (see e.g., \[614, 622–632\]). In the Type-II and Type-III seesaw models, there exist new gauge-charged scalars or fermions that can furthermore be the targets of collider searches \[633–640\].

Leptogenesis has also been considered in lepton number violating phase transitions. The effect of a second-order phase transition in low scale resonant leptogenesis was explored in \[641\]. Since the discovery of gravitational waves, several works have studied first-order lepton number violating phase transitions. The scenario of leptogenesis via a first order CP-violating phase transition was investigated in Ref. \[642\]. This differs from conventional high-scale leptogenesis scenarios as the physics UV-completing the Weinberg operator does not need to be specified, and such an approach permits mass model independence. A scenario similar to EWBG was explored in \[643\], where the first order phase transition occurs at the seesaw scale, and the dynamics of the RHNs must be accounted for. In \[644\], the effects of the additional departure from equilibrium due to first- and second-order phase transitions was studied, along with the suppression of the asymmetry from new interactions induced by the symmetry-breaking sector.

The possible connection between gravitational waves and leptogenesis has been studied in several works, including \[645\] which considered the network of local cosmic strings produced by the spontaneous breaking of \(U(1)_{B-L}\). Such a scenario predicts a large stochastic gravitational wave background and accommodates high-scale thermal leptogenesis. In \[646\], the authors highlighted that the stochastic gravitational waves from the cosmic string networks generic prediction of the seesaw mechanism, and such a signal could be a possible probe for thermal leptogenesis. The connection between gravitational waves produced from cosmic string decay and low scale leptogenesis has also been explored \[647\].

### 3.2. Electroweak baryogenesis

[Contributor(s): Djuna Croon, Oliver Gould, Jorinde van de Vis]

In EW baryogenesis (EWBG),\(^7\) the Sakharov conditions for the generation of the matter-antimatter asymmetry (baryon number violation, C- and CP-violation and out-of-equilibrium dynamics) are satisfied at the EW scale. The EWPhT is a first order process (occurring through bubble nucleation) due to new BSM degrees of freedom, typically an extended scalar

\(^7\) For reviews, see \[648–653\]
sector. The CPV in the SM’s CKM matrix is supplemented by additional CPV from the BSM sector. The CP asymmetry in EWBG is induced by the passage of the bubble walls through the plasma, causing a separation of particles, which is then converted into a baryon asymmetry through EW sphalerons. As a consequence of its relatively low energy scale, typical EWBG scenarios can be tested in experiments, such as colliders, B-factories, EDM measurements and gravitational wave telescopes. However, in order to use the results of experiments to draw a conclusion about the validity of EWBG in light of existing constraints, a number of theoretical challenges have yet to be overcome.

a. Outline of the computation of the baryon asymmetry

The two commonly used approaches to the computation of the baryon asymmetry are illustrated in Figure 4.

The most well-established approach is the semi-classical method \([654, 656]\), in which the quantum Boltzmann equations are obtained via a gradient expansion (GE), which relies on the typical de Broglie wavelengths of the particles being small compared to the bubble wall width, and an on-shell (OS) relation. The latter relates the Green’s functions to the particle distribution functions. The ‘standard’ (fermionic) CP-violating source term, the so-called semi-classical source, coming from the space-time dependence of the CP-violating mass-terms, arises at second order in derivatives. When the system allows for flavor oscillations, a source term can arise already at first order in derivatives \([657, 659]\). In addition to the CP-violating source, the Boltzmann equations typically contain several interaction rates which relax the CP-asymmetry. The quantum Boltzmann equations can either be solved numerically (NS), as in \([659]\), or brought to a simpler set of diffusion equations by expanding the distribution function around its equilibrium shape and taking moments of the Boltzmann equations (M). The latter approach requires the additional assumption that the flavor-off-diagonal densities are suppressed. The diffusion equations for EWBG with fermions displaying flavor oscillations have not yet been derived, but are expected to yield a larger asymmetry than the case with the semi-classical force term only.

An alternative, often used approach to derive the diffusion equations uses the so-called VEV-insertion approximation (VIA) \([660, 663]\), which relies on a simple relation between the self-energy and the Green’s function. The VEV-dependent part of the mass is treated as a perturbation on top of the thermal masses. Upon using the on-shell relation, a CP-violating source and a CP-conserving relaxation rate arise at first and zeroth orders in derivatives respectively. These terms enter the diffusion equations which are obtained from Fick’s law.
The success of EWBG depends sensitively on the equilibrium thermodynamics of the phase transition, as well as on the bubble nucleation rate. The transition must be of first order for EWBG to be possible at all, and it must be relatively strong in order to produce the observed baryon asymmetry. The final baryon asymmetry scales roughly as $\eta_B \sim 1/T^2$ [668], where $T$ is the nucleation temperature, neglecting the temperature dependence of the bubble wall speed.

Reliably computing the baryon asymmetry from a specific model scenario is technically
challenging, due in part to complicated plasma and nonequilibrium physics. The calculation of (near) equilibrium quantities is significantly simpler, allowing greater theoretical control. One can both reach higher perturbative orders and utilise direct lattice Monte-Carlo simulations. This has elucidated the structure and quantitative reliability of relevant weak-coupling expansions. The loop expansion must be resummed in order to correctly capture all contributions at a given power in coupling \cite{669,672}. This can be achieved by organising the perturbative expansion by energy scale, whereby the thermal scale gives the largest contributions, but bosonic modes at lower energy scales have stronger effective couplings. In the vicinity of a phase transition, the leading thermal-scale contributions to the change in free energy approximately cancel, and the contributions of more strongly coupled infrared scales grow in relative importance. As a consequence, the inclusion of one-loop corrections from the infrared modes is usually necessary even to predict the order of the phase transition. Further, it has been demonstrated that one must include two-loop thermal effects to achieve better than $O(1)$ numerical accuracy \cite{673,676}. The thermodynamic quantity with largest uncertainty is presently the bubble nucleation rate, as it involves both spacetime dependence and near equilibrium physics, though recent work has improved its determination \cite{677,681}.

c. Wall velocity

Knowledge of the wall velocity is not only essential for the computation of the baryon asymmetry, but also for the determination of the gravitational wave signal that gets sourced by the bubble collisions. For a long time, it was thought that the baryon asymmetry is strongly suppressed for fast-moving bubble walls, leading to the assumption that successful EWBG and an observable gravitational wave signal were mutually exclusive. However, in \cite{682} the diffusion equations were derived, with a modified assumption for the shape of the distribution functions. The conclusion of \cite{682} was that EWBG is possible with supersonic wall velocities; this result was confirmed by the analysis of \cite{683}, where additional moments of the Boltzmann equations were taken. The determination of the wall velocity requires a separate computation, which is model-dependent, since the friction forces depend on the particle content of the plasma. The computation of the wall velocity is a topic of very active debate. In the original works \cite{684,685}, the wall velocity was determined by a solution to the Boltzmann equations and relies on the fluid ansatz. In this approach, the friction is singular when the wall speed equals the speed of sound. In \cite{686} it was demonstrated that this ‘sonic boom’ is absent when the deviation from equilibrium is parameterized differently. This is in conflict with the results of Ref. \cite{687}, where it was argued that a discontinuity
at the sound speed generically arises as a result of energy-momentum conservation. In principle, a full numerical simulation could settle this debate. A first step towards such a solution was obtained in [688], but as the authors did not simulate the massless background, a conclusion about the existence of the sonic boom could not yet be drawn.

d. Perturbative interaction rates The transport equations governing the calculation of the CP-asymmetry involve a number of interaction rates, describing the scattering and decay processes which take place in the plasma; see Ref. [668] for a recent compilation. These are either functions of momenta if the Boltzmann equations are tackled directly, or constants (integrated over momenta) if reduced to diffusion equations. In the VIA approach the thermal decay widths (and the thermal masses) enter in the expression for the source. While sufficiently fast interaction rates may be treated as in equilibrium, and sufficiently slow ones neglected, in EWBG there are nevertheless a large number of relevant and comparable interaction rates, with the precise number depending on the specific BSM degrees of freedom. All relevant $1 \to 2$ and $2 \to 2$ perturbative interactions must be included to capture the correct behaviour at leading logarithmic accuracy [689], i.e., corrections suppressed by $O(1/\log g^{-1})$ for perturbative coupling $g^2$. Soft near-collinear $1 + N \to 2 + N$ splittings arise at the same power in $g^2$, being enhanced by the Landau-Pomeranchuk-Migdal (LPM) effect [690, 691], a collective effect of the plasma. Therefore a complete calculation to leading order in powers of the coupling requires the inclusion of all such interactions [692, 693]. Beyond this, subleading corrections are often suppressed by only one power of $g$, due to both kinematic and Bose enhancements [692, 694]. At present, EWBG calculations typically make use of interaction rates partially at leading logarithmic order, and partially at full leading order [668, 695]. The numerical importance of corrections to these rates is unclear, but may be sizeable, e.g., for the dilepton production rate, relevant to leptogenesis, a factor of $\sim 3$ enhancement was found to arise from the LPM effect [696, 697], and a 30-40% enhancement from $O(g)$ corrections [698]. Thus, calculating corrections to the interaction rates entering EWBG is an important challenge, likely necessary to resolve $O(1)$ numerical uncertainties.

e. Electroweak sphaleron rate EW sphaleron interactions in front of the bubble wall, biased by the C and CP asymmetries in the plasma, rapidly produce a net baryon number [406, 699, 701] (with rate $\Gamma_{sph} \sim 120 \alpha_w^5T \gg H$ [702, 704]). Successful EWBG depends on sphaleron transitions being shut off in the broken phase, as the sphaleron rate is proportional to $\Gamma_{sph} \sim \exp(-E_{sph}(T)/T)$ where $E_{sph}(T)$ is the sphaleron energy (which grows as $T$ falls).
An oft-employed estimate to check this is through the so-called washout avoidance condition 
\( \frac{v_\phi(T_C)}{T_C} \gtrsim 1 \) (where \( T_C \) is the critical temperature and \( v_\phi(T_C) \) the corresponding Higgs VEV). However, this is an estimate at best, plagued by different uncertainties. For example, it is not a gauge invariant quantity, and moreover, BSM effects may contribute to the sphaleron energy and therefore change the appropriate value of the numerical factor in the avoidance condition (e.g., \([705, 706]\)). A further approximation which is often made is to set the weak hypercharge to zero, \( g' \to 0 \), which corresponds to an error of \( \delta E_{sph}/E_{sph} \sim \sin^2 \theta_W \sim 10^{-2} \) \([707, 708]\). For non-zero \( g' \) or if custodial symmetry is broken (which happens in both the SM and BSM theories), the sphaleron configuration may deviate from SO(3) symmetry \([705]\). A complete and direct calculation of the sphaleron energy in the early-universe plasma is preferred (such as in \([709,711]\)), though troubled by the same thermodynamic uncertainty noted in the paragraphs above.

\[ f. \quad \text{Current and future constraints on model space} \]

Measurements at the LHC and constraints on the EDM of the electron \([712]\) put severe constraints on many models of EWBG, and they render the traditionally popular implementations via the MSSM \([713,716]\) and two-Higgs doublet model \([717,718]\) inviable \([653,719,721]\). The existence of stringent constraints on new CP violating sources from the EDMs underlines the importance of finding alternative models for EWBG and to provide accurate calculations of the resulting baryon asymmetry. Examples of mechanisms for EWBG that are consistent with experimental constraints are models with a temperature-dependent CKM-matrix \([722]\), CP-violation in a two-step phase transition \([723,724]\), composite dynamics \([725]\), symmetry non-restoration \([726]\) and cancelled \([727]\) or small \([728]\) EDMs (the latter two used the VIA approach to estimate the baryon asymmetry). Future gravitational wave experiments such as LISA may detect a stochastic gravitational wave background resulting from a first order EWPhT (e.g. \([729]\)), with phenomenology associated with the particular model under study. The bubble wall velocity is of crucial importance: in \([724,730,731]\) it was found that only detonation-like phase transitions produce an observable gravitational wave signal, while EWBG is typically associated with subsonic wall velocities.

### 3.3. WIMP baryogenesis

[Contributor(s): Yanou Cui, Michael Shamma]

WIMP triggered baryogenesis models are motivated by the fact that WIMP thermal
freeze-out satisfies Sakharov’s out of equilibrium condition needed for successful baryogenesis. The two main mechanisms of WIMP baryogenesis include: through WIMP annihilation around the freezeout time (e.g. [732–734]), and through the post-freezeout decays of a WIMP (e.g. [735–738]). These scenarios yield new opportunities for explaining the observed coincidence $\Omega_{\text{DM}} \sim 5\Omega_B$ between the DM and baryon abundances.

**WIMPy baryogenesis from WIMP DM annihilation.** This type of model is based on the subtle fact that during WIMP freeze-out, the net DM departure from equilibrium becomes significant at $T \sim m_{\text{DM}}$, which occurs before the usual freezeout temperature $T_{\text{fo}} \sim m_{\text{DM}}/O(10)$. The subsequent separation in time permits the build-up of baryon asymmetry through DM depletion before its freeze-out if the annihilation is CP- and B (or L)-violating. Washout processes suppress the produced baryon asymmetry, but these processes can slow down to below the Hubble rate before $T_f$. Solving the Boltzmann equations for the DM and baryon abundances, one finds the following approximation for the co-moving baryon asymmetry

$$Y_{\Delta B}(x \to \infty) \approx \frac{\epsilon}{2} [Y_{\text{DM}}(x_{\text{washout}}) - Y_{\text{DM}}(x \to \infty)]$$

where $Y_{\text{DM}}(x_{\text{washout}})$ is the co-moving DM density at the time of washout processes freezing out, $Y_{\text{DM}}(x \to \infty)$ is the observed co-moving DM density, and $\epsilon$ is the CP asymmetry factor which gives the net baryon asymmetry for each DM annihilation.

A particular realization of this scenario is that in which the WIMP annihilation violates lepton number and produces an asymmetry in leptons. The lepton asymmetry is converted into the observed baryon asymmetry by EW sphalerons. A Lagrangian realizing this scenario includes the following interactions:

$$\mathcal{L} \supset (\lambda_i X^2 + \lambda'_i \bar{X}^2) S_i + \lambda_{\psi_i} L\psi S_i$$

where DM consists of a gauge-singlet pair of Dirac fermions $X$ and $\bar{X}$ coupled to pseudo-scalar gauge singlets $S_{1,2}$. Additionally, the pseudo-scalar singlets couple to weak-scale $SU(2)_L$ doublet fermions $\psi$ and the left-handed SM lepton doublet. $X - \bar{X}$ annihilation to $L, \psi$ trigger baryogenesis. The requirement that washout processes become ineffective before $X$ freezeout can be realized for $m_\psi \gtrsim m_X$, while $m_\psi < 2m_X$ is necessary for $X$ annihilation to kinematically allowed. For further details see [732].
**Baryogenesis triggered by metastable WIMP decay.** In this type of model the baryon asymmetry arises from the CP- and B(L)-violating decays of a metastable WIMP parent $\chi$ after its annihilation freezes out. The abundance of $\chi$ at the time of its freeze-out sets the initial condition for baryogenesis. The late decay may occur in a wide time window prior to the BBN, while the predicted baryon asymmetry is not sensitive to the lifetime of $\chi$. The solution in the weak washout regime gives a simple relationship between the baryon asymmetry today and the $\chi$ abundance at the time of its freeze-out

$$Y_B(0) \approx \epsilon Y_{\chi_B}(T_{f.o.}), \quad \Omega_B(0) = \epsilon \frac{m_p}{m_{\chi_B}} \Omega_{\chi_B}^{\tau \to \infty}$$

where $\Omega_{\chi_B}^{\tau \to \infty}$ is the “would-be” abundance of $\chi$ in the limit that it is stable, $m_p$ is the proton mass and $\epsilon$ is the CP asymmetry factor. Therefore, a generalized WIMP miracle applies to baryon abundance. A Lagrangian realizing this scenario includes the following interactions

$$\Delta \mathcal{L} = \lambda_{ij} \phi d_i d_j + \alpha_i \chi_B \bar{u}_i \phi + \beta_i \psi \bar{u}_i \phi + \eta \chi_B^2 S + \gamma |H|^2 S + \text{h.c.}$$

where all couplings can be complex, $H$ is the SM Higgs boson; $d_i$ and $u_i$ are right-handed SM quarks with flavor indices $i = 1, 2, 3$; $\phi$ is a di-quark scalar with the same SM gauge charge as $u$; $\chi_B$ and $\psi$ are SM singlet Majorana fermions, and $S$ is a singlet scalar. The embedding of this mechanism in SUSY theories can be found in [735, 736].

Recent developments related to this scenario include, for instance, **WIMP Cogenesis** and **Dark Freezeout Cogenesis** [737, 738]. In the former [737] of these models, the decay of the WIMP grandparent simultaneously generates DM and baryon asymmetries. Simultaneous production gives no ambiguity in predicting the DM-baryon coincidence, permits a generalized baryon/lepton number symmetry $U(1)_{B(L)+DM}$, and provides a WIMP miracle abundance to ADM. As in conventional models of ADM, the relic DM abundance is composed of GeV-scale mass particles. Through renormalizable interactions with quarks, the DM candidate is within reach future iterations of DarkSide [739] and other upcoming direct detection experiments. In the latter [738] of these new models, the DM-baryon coincidence is achieved with number conserving interactions between stable and metastable DM partners in a hidden sector. Number conservation relates the partners’ abundances at their freeze-out. The metastable DM partner subsequently undergoes CP- and B-violating
decays and transfers its abundance into a baryon asymmetry which is related to the stable DM abundance through the aforementioned number conservation.

**Phenomenology.** The phenomenology predicted by WIMP baryogenesis mechanisms are rather rich, including various signals relevant for DM direct and indirect detection experiments, as well as intensity frontier experiments (e.g. EDM measurements). Additionally, these models may contain new particles with masses and interactions at or near the EW scale, and thus can be within the reach of the current or near future particle collider experiments. Notably, in the case of WIMP baryogenesis from decays, the WIMP parent must survive its thermal freeze-out time in order to meet Sakharov out-of-equilibrium condition, thus with a lifetime

$$\tau_{\text{WIMP}} \gtrsim \left( \frac{T_{\text{f.o.}}}{100 \text{ GeV}} \right) \times 10^{-10} \text{ sec}.$$  \hspace{1cm} (11)

This relatively long lifetime corresponds to a decay length of $l \sim 1 \text{ mm}$, which is intriguingly around the tracking resolution scale of detectors at collider experiments such as the LHC or future high luminosity collider experiments. As such, once the WIMP is produced, its subsequent decays would generate displaced vertex signatures \[740\]. WIMP baryogenesis has become a benchmark case for long-lived particle searches at current/planned collider experiments \[632, 741, 742\].

Additionally, there may be signals of induced proton decay \[242\]. For example, in the WIMP cogenesis scenario, the scattering of ADM with the proton effectively proceeds with a dimension-7 effective operator, and can be estimated as:

$$\sigma_{\text{IND}} \sim \frac{1}{16\pi^3} \left( \frac{\alpha^2 \beta \gamma \eta_2 m_p m_\chi}{m_{\text{WIMP}}^3} \right)^2$$

where $\alpha, \beta, \gamma, \eta$ are generic couplings of $O(1)$, $m_p$ is the proton mass, $m_\chi$ is the ADM mass, and $m_{\text{WIMP}}$ is the WIMP mass. This leads to a prediction for the proton lifetime as $\tau_p^{-1} = n_{DM}\sigma_{\text{IND}} v$. This model can lead to a proton lifetime that is consistent with the current lower bound set by Super-Kamiokande searches \[743\] while within reach of future experiments such as Hyper-Kamiokande \[744\] and DUNE \[745\].

### 3.4. Baryogenesis by particle-antiparticle oscillations

[Contributor(s): Seyda Ipek]

Pseudo-Dirac fermions can undergo particle–antiparticle oscillations similar to neutral meson oscillations. Due to the existence of both the Dirac and Majorana masses, the mass
eigenstates are a mixture of particle and antiparticle interaction states. Furthermore, if both the particle and the antiparticle are allowed to decay into the same final state, there can be $CP$ violation due to a physical phase difference in the respective coupling constants $[746]$. The oscillations can enhance $CP$ violation in the parameter regime where the mass difference between the heavy and light mass eigenstates ($\Delta m$) is the same order of magnitude as the decay width of the particles ($\Gamma$). Namely, there can be a large amount of $CP$ violation if the particle/antiparticle system oscillates a few times before they decay. Conversely, if the oscillations are too fast ($\Delta m \gg \Gamma$), the $CP$ violation is washed out and if they are too slow ($\Delta m \ll \Gamma$) decay happens before oscillations and $CP$ violation is again reduced.

In the early universe, the oscillation dynamics are affected by both the expansion of the universe and the interactions of the pseudo-Dirac fermions with the SM plasma. For example, oscillations do not start until the Hubble rate drops below the oscillation frequency, $H(T) < \omega_{\text{osc}} = \Delta m$. For a mass difference of $\sim 10^{-4}$ eV and smaller, this means that oscillations would be delayed until the temperature of the universe is smaller than the mass of the particles, $T < M$. Then, in order for the $CP$ violation in these oscillations to be important, the decays should also be delayed causing the particles to decay out of equilibrium. Assuming these decays are baryon-number (or lepton-number) violating, all three of the Sakharov conditions could be satisfied.

The interactions of the pseudo-Dirac fermions with the SM plasma can further hinder the oscillations in the early universe. These interactions can either be $flavor-blind$ or $flavor-sensitive$ depending on if the Lagrangian is symmetric or antisymmetric under $\psi \rightarrow \psi^c$, respectively. Here $\psi$ is the particle state and $\psi^c$ is the antiparticle state. If the pseudo-Dirac fermions have flavor-sensitive elastic scatterings with the SM plasma, oscillations are delayed until $\Gamma_{\text{scat}}^{\text{fs}} < \omega_{\text{osc}}$. Although flavor-blind elastic scatterings do not cause decoherence, flavor-blind annihilations still delay oscillations.

The Boltzmann equations that govern the time evolution of the density matrix $Y \equiv n/s \propto \sum_{\psi, \psi^c} |\psi_i\rangle\langle \psi_i|$ for particles $\psi$ and antiparticles $\psi^c$ are given as

$$zH \frac{dY}{dz} = -i(\mathbf{HY} - \mathbf{YH}^\dagger) - \frac{1}{2} \sum_{+, -} \Gamma_{\pm} [O_{\pm}, [O_{\pm}, Y]] - \sum_{+, -} s\langle \sigma v \rangle_{\pm} \left( \frac{1}{2} \{Y, O_{\pm} Y O_{\pm}\} - Y_{\text{eq}}^2 \right),$$

(12)

where $z = M/T$, $H$ is the Hubble rate and $s$ is the entropy density. The first term on the left-hand side describes the oscillations with the Hamiltonian $\mathbf{H} = M - i\mathbf{\Gamma}$. $\Gamma_+/\Gamma_-$
is the rate of inelastic scatterings for flavor-sensitive/blind interactions respectively, $\langle \sigma v \rangle$ is the thermally-averaged annihilation cross section and $O_{\pm} = \text{diag}(1, \pm 1)$ differentiates between flavor-sensitive and flavor-blind interactions. Note that the second term is zero for flavor-blind interactions.

These equations can be solved numerically for model-independent, generic models [747]. The resulting baryon asymmetry can be well approximated by

$$\Delta_B \simeq \epsilon \Sigma_\psi(z_{\text{osc}}) ,$$

where $z_{\text{osc}}$ is when oscillations start and $\Sigma_\psi = Y_\psi + Y_{\bar{\psi}}$. The $CP$-violation is quantified by the parameter $\epsilon = (\Gamma(\psi \rightarrow BX) - \Gamma(\psi \rightarrow \bar{B}X))/\Gamma$, where $X$ is a state with zero baryon number. Figure 5 shows the evolution of the $\psi$ number density and the creation of the baryon asymmetry for a set of parameters.

This mechanism can be easily realized in $U(1)_R$-symmetric MSSM (MRSSM), see e.g. [748]. In MRSSM, gauginos are necessarily pseudo-Dirac fermions. If there is also $R$-parity-violating interactions, then a pseudo-Dirac bino can go under $CP$-violating oscillations and $B$-violating decays. This scenario is realized for representative mass scales of $\mathcal{O}(100)$ GeV bino and $\mathcal{O}(10)$ TeV sfermions.

3.5. Mesogenesis

[Contributor(s): Gilly Elor, Robert McGehee]

Mesogenesis is a new, experimentally testable mechanism of low-scale baryogenesis and DM production which utilizes CPV in SM mesons [749–751]. In this mechanism, a scalar field $\Phi$ with a mass of $10−100$ GeV decays at a low temperature $T_R$ to $q\bar{q}$ pairs. $\Phi$ may or may not be related to inflation, but $T_{\text{BBN}} \lesssim T_R \lesssim T_{\text{QCD}}$, so there exists a late matter-dominated era. At such MeV scales, the $q\bar{q}$’s subsequently hadronize into SM neutral and charged mesons which undergo out-of-equilibrium CPV processes such as neutral $B^0_{d,s}$ oscillations or charged meson decays. These processes are expected in the SM, but CPV contributions from new physics could exist (and are required in some models of Mesogenesis). Baryon number is never violated thanks to the introduction of a new dark sector fermion $\psi_B$ carrying baryon number $B = -1$.

There are two sub-classes of Mesogenesis models. In the first, the daughter meson of the CPV process decays into the dark baryon and a SM baryon generating an equal and opposite
FIG. 5. The total $\psi$-number density $\psi(z)$, the $\psi$-asymmetry $\Delta(z)$ and the baryon asymmetry $\Delta B(z)$ for an effective interaction cross-section of $\sigma_0 = 1$ fb (solid), 1 ab (dashed), $10^{-2}$ ab (dotted) and $M = 300$ GeV, $\Delta m = 2m = 4 \times 10^{-6}$ eV and $\Gamma = 10^{-6}$ eV. $r = 0.1$ is a decay parameter and $\sin \phi_T = 0.5$ is a $CP$-violating phase. The baryon asymmetry of the Universe, $\eta \simeq 10^{-10}$, is shown for reference. The oscillations are delayed longer for flavor-sensitive interactions: For an effective cross section $\sigma_0 = 1$ ab (dashed) the oscillations start at $z_{\text{osc}} \sim 9$ if the interaction is flavor blind, while they start at $z_{\text{osc}} \sim 20$ if the interaction is flavor sensitive. (Figure is from [747].)

baryon asymmetry between the dark and visible sectors. Since the stability of matter requires $m_B \gtrsim m_p$, this is only possible for sufficiently heavy daughter mesons. In the second, the daughter meson decays instead into a pair of dark and SM leptons, generating an equal and opposite lepton asymmetry between the dark and visible sectors. This lepton asymmetry is then transferred to a baryon asymmetry between the two sectors via dark-sector processes. This second sub-class, while requiring extra dark-sector dynamics, allows the usage of CPV in lighter SM mesons to generate the baryon asymmetry. These Mesogenesis mechanisms are summarized in Table. II. We now abridge models of each sub-class starting with two models from the first.

In neutral $B$ Mesogenesis [749], the CPV of $B_{s,d}^0 - \bar{B}_{s,d}^0$ is leveraged. To mediate the decay into the dark state, one introduces a colored triplet scalar $Y$ with electric charge $-1/3$ and baryon number $-2/3$. The following interactions are allowed $\mathcal{L}_Y = - \sum_{i,j} y_{ij} Y^* \bar{u}_{i,R} d_{j,R}^c - \sum_k y_{\psi_B} Y \bar{\psi}_B d_{k,R}^c + \text{h.c.}$ Consistency with LHC bounds requires $M_Y \sim \mathcal{O}(\text{TeV})$, so integrating out this scalar yields the effective operator:

$$\mathcal{O} = \frac{y^2}{M_Y^2} \bar{u}_i^c d_j^c \bar{B} \psi_B + \text{h.c.}$$ (14)
where \( y^2 \equiv y_{ij}y_{i'j'} \). This allows the \( b \) quark within the meson to decay via \( \bar{b} \to \psi_Bud \). After undergoing oscillations, the \( B \) meson decays into dark and SM baryons, resulting in an equal and opposite baryon asymmetry between the dark and visible sectors. The baryon asymmetry is directly linked to experimental observables and successful Mesogenesis requires

\[
A_{sl} \times \text{Br} \left( B^0 \to B_{\text{SM}} + \psi_B \right) \gtrsim 10^{-7},
\]

where \( A_{sl} \) is the semi-leptonic asymmetry. Searches for the apparent baryon-number-violating meson decays in this mechanism are already underway at Belle \[752\] and LHCb \[753, 754\]. Furthermore, the UV model giving rise to Eq. (14) also predicts new decay modes for strange baryons \[755\] which are being searched for by BES \[756\]. Since neutral \( B \) Mesogenesis predicts the existence of a colored triplet mediator, collider and flavor observables can indirectly probe this mechanism. Given the plethora of signals and ongoing experimental searches, neutral \( B \) Mesogenesis is likely to be fully probed within the next 5-10 years \[757\]. Neutral \( B \) Mesogenesis can also be explicitly realized in a SUSY model with Dirac gauginos and an \( R \)-symmetry identified with baryon number \[758\].

In \( B_c^+ \) Mesogenesis \[751\], \( B_c^+ \) undergoes a CPV decay to \( B^+ \) which subsequently decays into the dark sector via the operator in Eq. (14):

\[
B_c^+ \to B^+ + f, \quad B^+ \to \psi_B + B^+.
\]

The baryon asymmetry is directly controlled by 1) the CPV in \( B_c^+ \) decays, 2) the branching fraction of the \( B_c^+ \) decay into \( B^+ \) mesons and other SM final states, and 3) the branching fraction of the \( B^+ \) meson into SM baryons and missing energy. The first observable is expected to be sizeable \[759\] but is currently not well-constrained, nor is the second. However, the branching fraction of \( B^+ \) is being probed by the same searches as neutral \( B \) Mesogenesis. Overall, this is a remarkably simple model of Mesogenesis and provides motivations for \( B_c \) physics searches at e.g. the LHCb \[760\] and an electron Future Circular Collider \[761\].

In the second sub-class of Mesogenesis models, the daughter mesons of the CPV process are too light to decay to a pair of dark and SM baryons. Instead, they decay into a pair of dark and SM leptons resulting in an equal and opposite lepton asymmetry between the dark and visible sectors. Two such models of Mesogenesis involve CPV decays of \( D^+ \) \[750\] and \( B^+ \) \[751\] mesons:

\[
D^+ \text{ or } B^+ \to M^+ + M, \quad M^+ \to \ell_d + \ell^+,
\]

(17a)
where $\mathcal{M}^+$ is a charged SM meson. Since this process occurs at MeV scales, EW sphalerons cannot convert this lepton asymmetry into a baryon asymmetry, but dark-sector scattering can sufficiently transfer the lepton asymmetry to a baryon asymmetry. As above, the generated lepton asymmetry is directly tied to experimental observables such as the CPV in a particular decay mode:

$$A_{\text{CP}} = \frac{\Gamma(D^+ \to f) - \Gamma(D^- \to \bar{f})}{\Gamma(D^+ \to f) + \Gamma(D^- \to \bar{f})}$$

(and the analogous definition for $B^+$ decays). To achieve a lepton asymmetry greater than the observed baryon asymmetry, the relevant CPV and branching ratios in each Mesogenesis model must satisfy

$$D^+ : \sum_{f \supset \pi^+} A_{\text{CP}}^f \text{Br}(D^+ \to f) \gtrsim 3 \times 10^{-5}, \quad \text{Br}(\pi^+ \to \ell_d + \ell^+) \gtrsim 10^{-3}, \quad (19a)$$

$$B^+ : \sum_{f \supset \mathcal{M}^+} A_{\text{CP}}^f \text{Br}(B^+ \to f) \gtrsim 5.4 \times 10^{-5}, \quad \sum_{\mathcal{M}^+} \text{Br}(\mathcal{M}^+ \to \ell_d + \ell^+) \gtrsim 10^{-3}. \quad (19b)$$

$D^+$ Mesogenesis may thus be probed by improved sensitivity to both CPV and branching ratios of $D^+$ decays to pions (at e.g. LHCb) and Br($\pi^+ \to \ell_d + \ell^+$). In $B^+$ Mesogenesis, it may be possible that the SM contains the necessary CPV and branching ratios required to produce the observed baryon asymmetry. Although it is difficult to calculate $A_{\text{CP}}^f$, some predicted branching fractions of $B^+$ are on the order of the current experimental central values \[762\]. It is instead easier to probe the decays of the lighter $\mathcal{M}^+$ to SM leptons + invisible (e.g. \[763\]–\[768\]), often by recasting searches for sterile neutrinos.

### 3.6. Axion baryogenesis

[Contributor(s): Raymond Co]

Axions are ubiquitous in particle physics as argued in Sec. \[2.9\] and we follow the notations and terminology defined therein. Here we review the development of axion baryogenesis and focus on how the axion dynamics in the early universe can play a role in explaining the observed baryon asymmetry of the Universe. We organize the discussion of the different mechanisms by the order of when the axion dynamics is potentially important, i.e., from during inflation to high and low temperatures after inflation.

During inflation, the effects of axion dynamics can be significant when the axion is the inflaton. For a standard cosine potential given in Eq. \[3\], the axion mass must be around
TABLE II. Summary of different flavors of Mesogenesis. We distinguish the four current mechanisms by the SM meson system whose CPV is leveraged, and the particle content of the dark sector. Also listed is the relevant experimental observables that the baryon asymmetry is linked to. Finally, we list the experiments relevant in searching for each mechanism. Note that model dependent and indirect signals are not shown but discussed in text.

$10^{13}$ GeV in order to successfully explain the cosmic perturbations. As a result, the axion discussed in this paragraph has to be an ALP rather than the QCD axion. If the ALP $a$ couples to the Chern–Simons term of the SM hypercharge gauge field $\gamma_{\mu
u}$,

$$\mathcal{L} \supset \frac{a}{4\Lambda} Y_{\mu\nu} \tilde{Y}^{\mu\nu}, \quad (20)$$

with $\Lambda$ is the suppression scale related to the axion decay constant, the non-zero ALP field velocity $\dot{a}$ leads to a helical hypermagnetic field via tachyonic instability [769–774], although an unnaturally large coupling constant is required when $\dot{a}$ originates from the axion oscillations. This long-range hypermagnetic helicity $\mathcal{H}_Y$ can remain until the EWPhT and even until today as the electromagnetic helicity. The field theoretic quantum anomaly of the hypercharge gauge group will induce a baryon asymmetry from the change of $\mathcal{H}_Y$ when the EW symmetry is unbroken. During the EWPhT, the hypermagnetic field is converted into the electromagnetic field, while the latter does not contribute to the baryon-number anomaly. Therefore, it is crucial to determine whether the conversion is completed only after the EW sphalerons go out of equilibrium so that a non-zero baryon asymmetry is generated. More generically, the decay of the hypermagnetic helicity is induced by the magnetohydrodynamic evolution as well as the EWSB. In the earlier works [775–782], mostly...
the latter is considered, while the former is not thoroughly taken into account; it is assumed that washout of baryon number by EW sphalerons is avoided, e.g., by a first order EWPhT. Refs. [783, 784] show that a non-zero baryon asymmetry remains even if there is no $B - L$ asymmetry and the EWPhT is a continuous crossover. This motivated the phenomenological study in Refs. [773, 783]; see also earlier works in Refs. [779, 781]. Ref. [786] further included the effects on the baryon asymmetry of chiral fermions that are necessarily produced along with hypermagnetic helicity during inflation [787, 788]. Moreover, in axion inflation, due to the chiral anomaly, the hypermagnetic helicity and baryon asymmetry are produced in a way that they can in principle annihilate each other, then while Ref. [786] showed that there is likely parameter space where the annihilation is incomplete and the observed baryon asymmetry is explained. To place emphasis on the particular relevance to this section, we note that, among the references above, Refs. [773, 779, 781, 785–787] specifically exploit both axion inflation and baryogenesis from decaying hypermagnetic helicity, while others concern one or the other. More recently, a mechanism named wash-in leptogenesis [789] pointed out that the $B - L$ asymmetry can be generated by $CP$ violation coming from the aforementioned chiral fermions with the aid of lepton-number-violating interactions of the right-handed neutrinos, so that the annihilation of the hypermagnetic helicity and baryon asymmetry is safely avoided. The additional effects of charged-lepton flavor violation are included in wash-in leptoflavorgenesis [790].

We now turn to the possibility that the axion is not the inflaton and evolves after inflation. Let us consider an axion field $a$ that couples to a current,

$$\mathcal{L} \supset \frac{\partial_\mu a}{f_a} J^\mu_Q \equiv \partial_\mu \theta J^\mu_Q \quad (21)$$

where $J^\mu_Q$ is a generic current to be specified below, $a$ can denote either the QCD axion or a generic ALP, and $f_a$ is the corresponding axion decay constant. When the axion field evolves coherently with a non-zero angular velocity $\dot{\theta} \equiv \dot{a}/f_a$, this coupling provides a bias between particles and antiparticles and may source baryon asymmetry. Refs. [791, 792] explore the possibility where $Q$ is the baryon number $B$ and some $B - L$ violation interaction is in thermal equilibrium. However, baryogenesis occurs in more generic situations [567, 793, 794]: even if $Q$ itself is not the baryon number, the SM interactions distribute the asymmetry to all particles, so any $Q$ and any baryon number violation can produce baryon asymmetry.

If the axion field velocity is driven by the vacuum potential, as in spontaneous baryoge-
nesis \cite{91, 92}, \( \dot{\theta} \simeq m_a \theta_i \) at the onset of the oscillation. Assuming that baryon violation is in thermal equilibrium, the baryon asymmetry is

\[
Y_B = \frac{n_B}{s} \sim \frac{1}{6} \frac{\dot{\theta} T^2}{s} \sim 10^{-10} \left( \frac{m_a \theta_i}{50 \text{ GeV}} \right)^{1/2}.
\]  

(22)

This clearly shows that the required axion mass is incompatible with the QCD axion and ALPs that are cosmologically stable to explain DM. In fact, the true baryon asymmetry production will be even less efficient because the magnitude \( \dot{\theta} \) redshifts from the initial value \( m_a \theta_i \) and \( \dot{\theta} \) alternates in sign throughout the oscillations. Note also that because of the large required axion mass, the production of baryon asymmetry occurs at a high temperature \( \sim 10^{10} \text{ GeV} \), and \( B - L \) violation is required.

As first considered in Refs. \cite{95, 96}, the angular velocity \( \dot{\theta} \) can be provided by a fast rotation in the angular direction of the field space of \( P \), where \( P \) is defined in Eq. (2) and contains the axion in the angular direction. A possible mechanism to initiate the rotation is provided by the Affleck-Dine mechanism and briefly described in Sec. 2.9. Refs. \cite{95, 96} consider the scenario where an interaction that simultaneously breaks both baryon symmetry and \( Q \) symmetry is in equilibrium, but as noted above, such a special interaction is not necessary \cite{567, 93, 94}.

For the QCD axion, the PQ asymmetry from the axion rotation generates the quark chiral asymmetry via the strong sphaleron process. A baryon asymmetry is then created from the quark chiral asymmetry by the EW sphaleron processes that fall out of equilibrium around the EWPhT at temperature \( T_{\text{EW}} \). If the QCD axion or an ALP has a weak anomaly, quark chiral and baryon asymmetries can also be directly generated from the EW sphaleron processes. Even though the PQ symmetry is violated by strong sphaleron, thermodynamic equilibrium retains most of the charge in the form of rotation to minimize free energy rather than in the form the chiral asymmetry \cite{567, 97}. Therefore, the rotation indeed survives to low temperatures, so that the final baryon asymmetry is determined at \( T_{\text{EW}} \) due to the termination of the charge transfer. Ref. \cite{567} names axiogenesis as the generic scheme where the PQ asymmetry is transferred to \( B + L \) by the QCD and EW sphaleron transitions (or \( B - L \) by other processes). Importantly, axiogenesis is qualitatively different from spontaneous baryogenesis in several points. 1) The velocity is driven by inertial motion, so the axion mass at the vacuum may be much smaller than that required in Eq. (22). Therefore, the QCD axion can produce the observed baryon asymmetry. 2) The rotation possesses a
(approximately) conserved charge so that the production of the baryon asymmetry may be understood as the flow of charge from the PQ charge to the baryon charge. As a result, no direct coupling between the axion and baryon current is required. 3) The (approximate) charge conservation and the rotation’s thermodynamic stability enable the production of baryon asymmetry down to low temperatures. In particular, the rotation may continue even after the EWPhT, so that a non-zero $B + L$ is protected against washout by EW sphaleron processes and therefore the minimal baryon number violation by EW sphaleron processes is enough for successful baryogenesis.

In fact, the thermodynamic stability of the rotation can also lead to peculiar observational signatures. The energy density of the rotation redshifts as matter at high temperatures and can come to dominate the total energy of the universe, while the scaling transitions to kinetic domination, which redshifts faster than radiation. If the universe is reheated by the rotation during these two phases, cosmic perturbations are created with an observable amount of non-Gaussianity \[798\]. The enhanced Hubble rate during these phases will enhance the gravitational waves from inflation and/or cosmic strings and leave a unique triangular spectrum \[799\] [801]. An additional gravitational wave signal may be generated by the rotation itself in the presence of a dark photon coupled to the axion \[802\] [803].

As discussed in Sec. 2.9, the axion rotation also contributes to the axion DM abundance via kinetic misalignment \[554\] [566]. To reproduce the observed abundances of DM and the baryon asymmetry, axiogenesis predicts a $T_{EW}$ at least a factor of 10 higher than the SM prediction since a lower $T_{EW}$ results in axion DM overproduction from kinetic misalignment.

This picture of axiogenesis is qualitatively different in the presence of $B - L$ violation at high temperatures since the EW sphaleron processes only violate $B + L$ and preserve $B - L$. This allows for production of $B - L$ at high temperatures and therefore the baryon asymmetry is dominantly produced at some other temperature higher than $T_{EW}$, which enhances the charge transfer efficiency. Examples of implementing such $B - L$ violation in the axiogenesis scheme are outlined in the following.\[8\] The dimension-5 Majorana neutrino mass terms can provide lepton number violation \[793\] [804] [805]. Dirac neutrino masses in composite theories can also produce effective $B - L$ asymmetry \[806\]. Baryon and/or lepton number violation

\[8\] Most recently, a mechanism called spontaneous leptoflavorgenesis \[790\] enables a non-zero baryon asymmetry to be produced from a large lepton flavor asymmetry even if the total $B - L$ vanishes. Although the large lepton flavor asymmetry may in principle result from an axion rotation, a rigorous study has not yet been carried out to show viable parameter space.
can arise from $R$-parity violation \cite{807}. An extra non-Abelian gauge symmetry $SU(2)_R$ can also give rise to baryon number violation \cite{808}. Since these axiogenesis scenarios enhance the charge transfer rate, axion DM is viable even with the SM prediction of $T_{EW}$. These scenarios prefer an axion decay constant $f_a = 10^{8-11}$ GeV, which is more experimentally accessible than the prediction from the conventional misalignment mechanism. Moreover, the simultaneous explanation of axion DM and baryon asymmetry by the axion rotation often leads to testable predictions.

4. OBSERVATIONAL ANOMALIES AND COSMOLOGICAL MODEL BUILDING

Over the past couple of decades, many experimental anomalies have come and gone. At face value, we therefore expect that any currently outstanding anomaly might have only a small chance of being a smoking-gun signature of the theory of physics beyond the SM. That being said, observational anomalies of the past, and attempts at reconciling them with theories of the early universe, have had a long-lasting impact in shaping what the theory community now considers to be part of the standard model-building toolkit. Hence, at the very least, we expect that the act of seriously engaging with anomalies in current observational data will bear fruit in stretching the boundaries of the theory landscape in the coming decade. Of course, any of these observations may also be the beginning of a revolution in our understanding of fundamental physics.

In the sections below, we outline several currently outstanding anomalies in cosmological and astrophysical data along with viable explanations involving theories of new physics. Sec. 4.1 begins with a discussion of the local Hubble expansion parameter and the amplitude of matter density fluctuations on large scales, the so-called $H_0$ and $\sigma_8$ tension, respectively. Sec. 4.2 discusses the physics of 21 cm cosmology, with a particular emphasis on the recent EDGES anomaly. Finally, Sec. 4.3 summarizes the current status in the disagreement between the predicted and inferred value of the primordial lithium abundance, while Sec. 4.4 discusses artifacts in observations of small scale structure within the context of self-interacting DM.
4.1. $H_0$ and $\sigma_8$ tension

[Contributor(s): Vivian Poulin]

The “Hubble tension” refers to the inconsistency between local measurements of the current expansion rate of the Universe, i.e., the Hubble constant $H_0$, and the value inferred from early-universe data using the $\Lambda$CDM model. This tension is predominantly driven by the Planck collaboration’s observation of the CMB, which predicts a value in $\Lambda$CDM of $H_0 = (67.27 \pm 0.60) \text{ km/s/Mpc}$ \cite{Planck}, and the value measured by the SH0ES collaboration using the Cepheid-calibrated cosmic distance ladder, whose latest measurement yields $H_0 = 73.04 \pm 1.04 \text{ km/s/Mpc}$ \cite{SH0ES}. Taken at face value, these observations alone result in a 5$\sigma$ tension.

Today, there exist a variety of different techniques for calibrating $\Lambda$CDM at high-redshifts and subsequently inferring the value of $H_0$, which do not involve Planck data. For instance, one can use alternative CMB data sets such as WMAP, ACT, or SPT, or even remove observations of the CMB altogether and combine measurements of BBN with data from BAO \cite{BAO1, BAO2} or with supernovae constraints \cite{Supernovae1, Supernovae2, Supernovae3}, resulting in $H_0$ values in good agreement with Planck. Similarly, alternative methods for measuring the local expansion rate have been proposed in the literature, in an attempt at removing any bias introduced from Cepheid and/or SN1a observations. The Chicago-Carnegie Hubble program (CCHP), which calibrates SNIa using the tip of the red giant branch (TRGB), obtained a value of $H_0 = (69.8 \pm 0.6 \text{ (stat)} \pm 1.6 \text{ (sys)}) \text{ km/s/Mpc}$ \cite{CCHP}, in between the Planck CMB prediction and the SH0ES calibration measurement. The SH0ES team on the other hand, using the parallax measurement of $\omega$–Centauri from GAIA DR3 to calibrate the TRGB, obtained $H_0 = (72.1 \pm 2.0) \text{ km/s/Mpc}$ \cite{SH0ES2, SH0ES3}. Additional methods intended to calibrate SNIa at large distances include: surface brightness fluctuations of galaxies \cite{SurfaceBrightness}, MIRAS \cite{MIRAS}, or the Baryonic Tully Fisher relation \cite{BaryonicTully}. There also exists a variety of observations which do not rely on observations of SNIa – these include e.g. time-delay of strongly lensed quasars \cite{Quasars1, Quasars2}, maser distances \cite{Maser}, or gravitational waves as “standard sirens” \cite{GravitationalWaves}. While not all measurements are in tension with Planck, these direct probes tend to yield values of $H_0$ systematically larger than the value inferred by Planck. Depending on how one chooses to combine the various measurements, the tension oscillates between the 4$\sigma$ and 6$\sigma$ level \cite{SH0ES, CCHP2, SH0ES4, CCHP3}.

On top of the statistically strong Hubble tension, the value of a quantity measuring
the amplitude of density fluctuations on large scales, namely $S_8 \equiv \sigma_8 (\Omega_m/0.3)^{0.5}$, inferred from CMB is about $2 - 3\sigma$ larger than that deduced from weak lensing surveys such as the CFHTLenS [833], KiDS, as well as from Planck SZ cluster abundances [809, 834]. In fact, the combination of KiDS data with BOSS and 2dFLenS data leads to a $3\sigma$ lower value than Planck, and points to a $\sigma_8$-tension rather than $\Omega_m$ [835]. However, some surveys such as DES [836] and HSC [837], while yielding $S_8$ lower than Planck, are statistically compatible with it at less than $2\sigma$. More data are therefore awaited to firmly confirm this potential breakdown of the $\Lambda$CDM model.

The most-precise parameter measured within CMB data, that drives constraints to $H_0$, is the so called angular scale of the sound horizon $\theta_s(z_*) \equiv r_s(z_*)/d_A(z_*)$, where $r_s(z_*)$ is the physical size of the sound horizon at recombination ($z_* \simeq 1100$) and $d_A(z_*)$ is the angular diameter distance to recombination. A naive increase of $H_0$ would decrease $d_A \propto 1/H_0$ (making the CMB appears closer to us) and therefore the typical size over which CMB exhibits fluctuation (given by $\theta_s$) would appear larger. This suggests two main ways of resolving the Hubble tension through new physics, based on the requirement to keep the key angular scale $\theta_s$ fixed, usually called late- and early-universe solutions (see e.g. [838, 839] for reviews). The first way exploits the geometric degeneracy that exists in CMB data: any solution that changes $H(z)$ at late-times ($z < z_*$) while keeping the angular diameter distance $d_A$ fix will leave the CMB mostly unaffected. As a result, it is possible to compensate a higher $H_0$ by decreasing $H(z > 0)$ (but before $z_*$), such that $d_A(z_*)$ and $r_s(z_*)$ are kept fixed. Moreover, assuming that the physical density $\omega_m$ is fixed by early-universe physics, any increase in $H_0 \equiv 100h$ necessarily leads to a decrease in $\Omega_m = \omega_m/h^2$, and therefore in $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$, providing a potential simple framework to resolve both tensions. The most popular example of such scenario is that of a phantom dark energy [840] with equation of state $w < -1$. There exists naturally a plethora of alternative proposals leading to more complicated behavior at late-times (see [831, 832, 841] for exhaustive reviews of models and application to the Hubble tension), including model-independent reconstruction of late-time dark energy [842, 843]. However late-universe models cannot simultaneously accommodate BAO data [847, 849] and SN1a from the Pantheon(+) catalogue [850, 851], while respecting the calibration from CMB data (measuring the sound horizon) and that from the local distance ladder (measuring the intrinsic SN1a luminosity) [839, 852, 855].
Therefore, the most promising way to resolve the Hubble tension amounts in compensating a smaller $d_A(z_*)$ due to an increase in the value of $H_0$ by reducing $r_s(z_*)$ (by $\sim 10$ Mpc [838, 856–859]) in the early-universe so as to keep $\theta_s$ fixed. Unfortunately, a generic feature of early-universe models suggested to resolve the Hubble tension is to increase the $S_8$ tension. This is because, to compensate the impact of the increased expansion rate onto the time-evolution of the potential well (the so-called ‘early integrated sachs wolfe effect’), which manifests as an increase in the height of the first acoustic peak, the DM density must increase [860, 861]. A similar increase in the DM density is required to keep the angular scale of the BAO fixed, such that it has been suggested that it might simply be impossible to resolve both tensions simultaneously with a single new-physics mechanism [861–864]. Bar-ring unknown systematic errors, more work needs to be done to find a satisfying common resolution to both tensions. Nevertheless, the future “concordance cosmology” will likely be based on some of the currently proposed solutions, and a great deal of information can be extracted from the current best set of solutions [839], that we summarize below.

- **New light relics and neutrino properties:** The first popular category of models makes use of new light degrees of freedom (similar to neutrinos), relativistic before recombination and typically parameterized by the effective number of relativistic degrees of freedom $N_{\text{eff}}$. Current constraints on additional free streaming species from Planck data and BAO data are very strong, restricting $N_{\text{eff}} < 3.33$ at 95% C.L [809] and therefore excluding an easy solution to the Hubble tension (which would require $N_{\text{eff}} \sim 4$). Interestingly, it is possible to completely change this constraint by relaxing the assumption of free-streaming. In fact, CMB data are powerful to test new secret interactions of neutrinos since these interactions affect the streaming properties of neutrinos [865, 871]. As a striking example, Planck 2015 temperature data can fully accommodate 4 species of strongly interacting neutrinos through a contact interaction, with $H_0 = 72.3 \pm 1.4$ km/s/Mpc [869, 872, 875]. Unfortunately, the simplest realization of this model is strongly constrained [876, 877] by including the newest Planck data [839, 878, 879] but such degeneracy provides 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 [867, 868, 880, 883], while potentially shedding light on the neutrino mass mechanism and leptogenesis [883]. There exists also a large number of scenarios invoking DR

---

9 Some authors prefer referring to it as the “least unlikely” category of solution [838].
beyond neutrino-like particles (e.g. thermal axions [884], DR from decaying DM [885–888], interfering DR [871, 878, 889], evaporating PBHs [890] – see review [832]). Additionally, models involving interacting dark radiation produced around matter-radiation equality (for instance following the decoupling of heavy particles from the thermal bath) can perform significantly better (and in fact are even favored by Planck data) than standard interacting dark radiation scenarios (present from early times), as shown recently in Ref. [891] for a simple realization of the Wess-Zumino supersymmetric model with the mass of the scalar mediator at the eV scale. Future CMB missions are expected to deliver high precision measurements of $N_{\text{eff}}$. In particular, the upcoming Simons Observatory [892] is expected to constraint $\Delta N_{\text{eff}} < 0.12$ at 95% CL, and proposed experiments such as CMB-S4 [893] could constraint $\Delta N_{\text{eff}} < 0.06$ at 95% CL.

• Early Dark Energy: Another popular way to decrease the sound horizon is to postulate the existence of an Early Dark Energy (EDE) component (often modelled as a scalar field) that behaves like a cosmological constant for $z \geq 3000$ and decays away as radiation or faster at later times [860, 894]. While many papers have investigated a modified axion potential [860, 894–898], a number of other models have been proposed (although not all models have the same level of success). These includes coupling the EDE scalar to neutrinos [899, 900] and to DM [901]; the New Early Dark Energy scenario where a first-order phase transition in the EDE field is triggered by another (subdominant) scalar field [902–904]; an EDE model with an anti-de Sitter phase around recombination [905]; a power-law potential with canonical [906] and non-canonical kinetic term [907]; an ALP sourcing DR [908]; an axion field tunneling through a chain of metastable minima [909]; models with $\alpha$-attractors behavior [910]. EDEs with scaling solutions in the matter- and radiation-dominated eras [911–914] are not capable of relieving the $H_0$ tension in a significant way [846, 915]. For model-independent reconstructions of the EDE fraction see Refs. [846, 916, 917]. The combination of Planck, BAO, SN1a and SH0ES data favor a model representing up to $\sim 10–15\%$ of the total energy budget in the redshift range $z \simeq 2000–5000$ (the exact redshift depends on the model) and diluting faster than radiation afterwards. However, it is unclear that Planck data alone favor such a model (see Refs. [897, 918] for discussion), although the EDE model resolving the tension does not degrade the fit to Planck data. Interestingly, the latest data from the Atacama Cosmology Telescope (ACT) favor EDE over $\Lambda$CDM at $\sim 2–3\sigma$ independently of any knowledge on the local $H_0$ value [917–919, 920], but the inclusion of
Planck high-\(\ell\) temperature data restricts that preference to below the 2\(\sigma\) level. The inclusion of Planck polarization data strengthen the preference \[921\], while SPT-3G data play little role in constraining the EDE model \[921, 923\]. Additionally, EDE models tend to slightly increase the \(S_8\) tensions \[863, 924\] and can be constrained with the full shape of the galaxy power spectrum \[925, 926\], although current data are inconclusive \[897, 927–929\]. Nevertheless, it is clear that EDE alone cannot resolve the \(S_8\) tension, and would require a more complicated dynamics \[901, 930, 931\] or an independent mechanism \[862–864\]. Future data will play a crucial role in arbitrating the fate of the EDE scenario, as it has been shown that data from the AdvACT \[919\] or CMB-S4 \[896\] can unambiguously confirm (or exclude) the presence of an EDE that resolves the Hubble tension in the early-universe.

- **Exotic Recombination:** An alternative way to reduce the sound horizon is by changing recombination physics, in order to increase the redshift of recombination \[932\]. A promising scenario leading to such effect is that of primordial magnetic field induced baryon clumping on small scales, where the presence of a primordial magnetic field of \(\sim 0.1\) nano-Gauss strength can generate baryon inhomogeneities on \(\sim \text{kpc}\) scales \[933, 934\]. Since the recombination rate is proportional to the electron density squared \(n_e^2\), the mean free electron fraction at a given epoch will be modified in the presence of an inhomogeneous plasma. This type of clumpy plasma recombines earlier, and thus reduces the sound horizon \(r_s\) at recombination. The corresponding effect on the CMB spectra is a shift in the position of the peaks, which can be compensated by a larger value of \(H_0\), without significantly spoiling the fit to CMB data \[935\]. It is also worth noting that the field strength required to solve the Hubble tension is of the right order to explain the existence of large-scale magnetic fields (without relying on dynamo amplification) and also slightly relieves the \(\sigma_8\) tension. However, such models are disfavored by BAO data \[862\]. The combination of ACT, SPT, Planck and BAO with a prior from SH0ES only leads to \(H_0 \simeq 69.28 \pm 0.56 \text{ km/s/Mpc}\) and therefore a mild alleviation of the Hubble tension \[936, 937\].

Alternatively, variation in fundamental constants \[938–942\] and in the CMB temperature \[943, 944\] have been shown to provide an interesting avenue to resolving the Hubble tension. One of the most successful solution involves a varying effective electron rest mass. When analyzing Planck and BAO data, the model leads to \(H_0 = (68.7 \pm 1.2) \text{ km/s/Mpc}\) \[940, 941\] with \(\delta m_e/m_e \sim 1\%\). Rather surprisingly, the inclusion of a small negative curvature \(\Omega_k \sim -0.01\) allows the model to reach \(H_0 = 72.3 \pm 2.8 \text{ km/s/Mpc}\) when confronted with the same
data set, without including information from the local $H_0$ measurements [941]. However, the inclusion of CMB lensing measurement from Planck, strongly affected by curvature [945, 946], restricts values of $H_0$ that can be reached to $H_0 = (69.3 \pm 2.1) \text{ km/s/Mpc}$ [839]. Going beyond, Ref. [316] suggests a solution that invokes a symmetry that rescales simultaneously the gravitational constant, the Thomson scattering rate and the primordial power spectrum amplitude, potentially realized through a mirror world made of dark atoms, with dark copy of the photons, baryons, and neutrinos [165, 316]. Such models could resolve both $H_0$ and $S_8$ tension, but requires a smaller primordial Helium fraction, at odds with current measurements.

**Modified Gravity:** Many scenarios invoking modification of gravity have been suggested to resolve either or both tensions [910, 947–965]. For instance the framework of scalar-tensor theories of gravity (Horndeski Theories [966]), in which a scalar field non-minimally coupled (NMC) to gravity dictates the strength of the gravitational field $G$, is particularly promising to resolve the $H_0$ tension. The simplest scalar-tensor theories of gravity can be written as

$$S = \int d^4x \sqrt{-g} \left[ \frac{F(\sigma)}{2} R - \frac{g^{\mu\nu}}{2} \partial_\mu \sigma \partial_\nu \sigma - V(\sigma) + L_m \right], \quad (23)$$

where $F(\sigma)$ is a generic function of the field $\sigma$ that controls the coupling to the Ricci scalar $R$, while $V(\sigma)$ describe the field’s potential. Constraints on such models are highly dependent on the form of $F(\sigma)$, and partly influenced by the choice of $V(\sigma)$. A successful proposal is for instance that of ‘early modified gravity’, an extension of the historical Brans-Dicke model [960, 961, 967], where $F(\sigma) = N^2_{\text{Pl}} + \xi \sigma^2$ and $V(\sigma) = \lambda \sigma^4$. In a similar way to EDE models, the scalar field is initially frozen deep in the radiation era, and when its effective mass becomes larger than the Hubble rate, it starts to perform damped oscillations about its minimum. However, due to the NMC parameter $\xi \geq 0$, the scalar field experiences a temporary growth before rolling down the potential, producing new features in the shape of the energy injection. In addition, the NMC predicts a weaker gravitational strength at early times, which leads to a suppression in the matter power spectrum at small scales (recalling that the growth of the matter perturbation $\delta$ is governed by $\ddot{\delta} + 2H\dot{\delta} = 4\pi G_{\text{eff}} \rho_m \delta$ well within the horizon). In Ref. [951], it was shown that the extra freedom introduced by $\xi$ allows the model to substantially relax the $H_0$ tension, even when Large Scale Structure data are included in the fit in addition to CMB and supernovae data. Furthermore, thanks to the fast rolling of $\sigma$ towards the minimum, the tight constraints on the effective Newton-
nian constant from laboratory experiments and on post-Newtonian parameters from Solar System measurements are automatically satisfied. Alternatively, studies involving torsional (or teleparallel) gravity \[956, 958, 968, 969\] as well as model-independent reconstruction of gravity \[844, 955, 957, 962, 965, 970, 972\] have shown promises in resolving cosmic tensions.

- **New DM properties:** The existence of new physics in the DM sector can easily help resolving the \(S_8\) tension by suppressing power on scales of order \(k \sim 0.1 – 1\) Mpc/\(h\). This can be done through DM decays \[886, 887, 973\], annihilations \[974\], or new interactions with neutrinos \[975, 979\], dark radiation (DR) \[878, 980, 988\], or dark energy \[989–991\]. Models of DM-DR interactions have been extensively studied in the context of the \(H_0\) and \(S_8\) tension \((e.g. \[980–984\]) because it was thought that they would provide an easy resolution to both tensions: indeed, the addition of new DR species would contribute to increase \(N_{\text{eff}}\), thereby resolving the \(H_0\) tension, while DM-DR interactions would suppress the growth of structure and accommodate a low \(S_8\). In practice, these models are very powerful for \(S_8\) but none can successfully explain \(H_0\) measurements, at best leading to a \(\sim 3\sigma\) residual tension \[839\]. Alternatively, it was originally suggested that the existence of a non-zero neutrino mass could resolve the \(S_8\) tension \[895, 992\], but current constraints on the neutrino mass exclude that possibility \[809, 973\], even in non-thermal scenarios \[993\]. Similarly, constraints on warm and fuzzy matter from Ly-\(\alpha\) data exclude a straightforward solution to the \(S_8\) tension \[994, 995\] unless they represent a small fraction of the total DM \[864, 996–998\].

One way of producing such a component at late-times is through DM decays into massive and massless daughters. Such models were introduced as potential solutions to the Hubble tension \[886, 888, 999\], but it was subsequently shown that CMB lensing, BAO and SN1a exclude these models as a resolution to the \(H_0\) tension \[885, 1000, 1004\]. However, models of DM decaying into a lighter state can explain the \(S_8\) tension, by producing a warm DM component through the decay of CDM (with data favoring a lifetime of \(\sim 50\) Gyrs), leading to suppressed clustering on scales probed by weak lensing experiments, in a similar fashion to massive neutrino or standard warm DM. Contrarily to the latter scenarios, the specific time dependence of the power suppression imprinted by the decay allows to accommodate simultaneously all data (and in particular CMB and lyman-\(\alpha\) constraints) \[973, 1005, 1006\].

Future experiments measuring the growth rate of fluctuations at \(0 \lesssim z \lesssim 1\) or the CMB lensing power spectrum to \(\sim 1\%\) accuracy will further test this scenario, and more generally solutions to the \(S_8\) tension.
4.2. 21 cm

[Contributor(s): Nadav Outmezguine]

The 21-cm signal is the result of the hyper-fine splitting of the hydrogen atom ground state. The energy difference between the singlet and the triplet states is of $\simeq 6 \times 10^{-6} \text{eV}$, corresponding to a wavelength of 21 cm. As hydrogen is the most prevalent element in the Universe, measurements of the (redshifted) 21-cm absorption line are commonly thought of as a probe of the baryonic gas in different epochs of the Universe [1007–1019]. In what follows we shall introduce briefly the important ingredients when studying the 21-cm absorption line (for a more detailed treatment see, e.g., [1015, 1016, 1019, 1020]) We then present a selection of observational efforts to measure the global 21-cm absorption line. We conclude by discussing the opportunities for observing signals of BSM particle physics in 21-cm cosmology, including, but not limited to, the “EDGES anomaly”.

The discussion in this section is restricted to the global (sky-averaged) 21-cm signal, though much like other cosmological observables, a lot of information is imprinted in the spatial fluctuations of the signal. We refer the reader to, for example, [1012, 1021, 1022] for more information, and to [1023] for an example use of 21-cm fluctuations to the study BSM physics.

4.2.1. Basics

The intensity of 21-cm photons is sensitive to the relative occupation of singlet and triplet states, most commonly quantified using the spin temperature, $T_s$, defined as

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-T_\star/T_s} = 3 e^{-T_\star/T_s},$$

(24)

where $n_1(0)$ is the triplet (singlet) number density, $g_1(0)$ is the triplet (singlet) statistical weight and $T_\star = 0.68 \text{K}$ is the energy difference between the states (the hyper-fine splitting energy). Importantly, the spin temperature does not measure the temperature of any thermodynamic system and is strictly defined by the relation above.

Three main processes effect the relative abundance of triplet and singlet states, and therefore govern the evolution of the spin temperature: (i) Absorption and emission of CMB photons. (ii) $H - H$, $H - e$ and $H - p$ collisions. (iii) Resonant scattering (to an off-shell excited state) of Ly$\alpha$ photons (the [1024, 1025]). Depending on which process dominates,
the spin temperature traces a different temperature and to a very good approximation it is given by \[1015\]

\[
T_s^{-1} = \frac{T_{\gamma}^{-1} + x_c T_K^{-1} + x_{\alpha} T_c^{-1}}{1 + x_c + x_{\alpha}}.
\] (25)

Above \(T_{\gamma}\) is the CMB temperature, \(T_K\) is the kinetic temperature of baryons and \(T_c\) is the color temperature of Ly\(\alpha\) photons\(^{10}\). \(x_c\) and \(x_{\alpha}\) are the collisional and Lyman – \(\alpha\) coupling coefficients, quantifying the strength of coupling of the spin temperature to the kinetic or color temperature, normalized to strength of coupling to CMB photons (see, for example Ref. \[1016, 1020\] for details on those couplings).

While CMB photons travel through a cold hydrogen gas background, 21-cm photons get absorbed and emitted, and then redshift until measured today. The relative abundance of hydrogen singlets and triplets at redshift \(z\) therefore affects the intensity of photons measured today with frequency \(\nu = \nu_0 / (1 + z)\), where \(\nu_0 \approx 1420\) MHz is the frequency of 21-cm photons. The (differential and co-moving) brightness temperature is then used as a proxy to the photons intensity, and is defined as

\[
T_{21}(z) = T_s(z) - T_{\gamma}(z) \frac{1}{1 + z} (1 - e^{-\tau(z)})
\]

\[
\simeq 27\text{mK} \ x_{\text{HI}}(z) (1 + \delta(z)) \left(1 - \frac{T_{\gamma}(z)}{T_s(z)}\right) \left(1 + \frac{z}{10}\right)^{1/2} \left(\frac{0.15}{\Omega_m h^2}\right)^{1/2} \left(\frac{0.023}{1\text{K}}\right) \frac{\partial_r v_r(z)}{(1 + z) H(z)}.
\] (26)

Above \(\tau(z)\) is the optical depth (survival probability) for 21-cm photons emitted at redshift \(z\). The second line assumes the optical depth is much smaller than 1; it is given generally by

\[
\tau(z) \simeq 9.2 \times 10^{-3} (1 + \delta(z)) x_{\text{HI}}(z) (1 + z)^{3/2} \left(\frac{1\text{K}}{T_s}\right) \frac{\partial_r v_r(z)}{(1 + z) H(z)},
\] (27)

which justifies this assumption. Above \(x_{\text{HI}}\) is the fraction of neutral (not ionized) hydrogen atoms, \(\delta = \delta \rho / \rho\) is the matter density contrast, \(H\) is the Hubble parameter and \(\partial_r v_r\) is the gradient of peculiar gas velocity along the line of sight. It is evident from Eq. (26) that \(T_{21}\) could only be observed when the spin temperature departs from the CMB. Then, a negative \(T_{21}\) corresponds to absorption of CMB photons, while a positive \(T_{21}\) corresponds to emission.

\(^{10}\) Ly\(\alpha\) photons do not follow a thermal distribution and therefore a Ly\(\alpha\) temperature is not well defined.

The color temperature is then defined in analogy with the temperature of a thermal distribution via \[1026\]

\[
T_c^{-1}(\omega) = -d \log f_{\text{Ly} \alpha} / d \omega, \text{ where } f_{\text{Ly} \alpha} \text{ is the density of Ly} \alpha \text{ photons}. \text{ Note that this temperature is a function of the frequency and not a constant.} \]
FIG. 6. The ΛCDM prediction for the cosmological evolution of the brightness temperature $T_{21}$. The black line represents an exemplar model that demonstrate the different phases of evolution of the brightness temperature as described in the main text, adopted from [1015]. The gray region represents the astrophysical uncertainties of the brightness temperature that appear once stars begin to form, as estimated in [1027]. Note that the region collapses into a line at earlier cosmic times, before stars ignite, demonstrating the certainty of the Dark Age prediction as opposed to the Cosmic Dawn and Reionization.

a. Cosmological context Having defined the brightness temperature, we next wish to understand how it cosmologically evolves. The Dark Age of the universe is the era between last scattering (CMB, $z \sim 1100$) and the onset of star formation, known as the Cosmic Dawn ($15 \lesssim z \lesssim 35$). The brightness temperature evolution could crudely be divided into a cosmology-driven phase during the Dark Age, and an astrophysics-driven phase during the Cosmic Dawn (and Reionization). Accordingly, the Cosmic Dawn evolution is much more uncertain than during the Dark Age. During the Dark Age, recombination occurs and only a small fraction ($\sim 10^{-4}$) of electrons remains free (unbound to protons). This small residual fraction is sufficient to maintain thermal equilibrium between CMB photons and the baryonic gas down to $z \sim 200$ [1016]. This means that in the redshift interval $z \gtrsim 200$, we expect $T_K \simeq T_\gamma$. Since star formation had not yet started, there is not any source of Lyman -- $\alpha$ photons, and by inspection of Eq. (25) we conclude that for $z \gtrsim 200$, $T_s \simeq T_\gamma$ and the brightness temperature vanishes, $T_{21} = 0$. After kinetic decoupling ($z \lesssim 200$) but yet before the Cosmic Dawn ($z \gtrsim 30$), the gas cools adiabatically and becomes colder than CMB photons $T_K < T_\gamma$. At the same time the collisional coupling dominates $x_c \gg 1$ and the
spin temperature traces the gas temperature, leading to a negative brightness temperature. At $z \sim 40$, adiabatic dilution of the gas renders the collisional coupling ineffective $x_c \ll 1$ and the spin temperature again follows the CMB, leading to $T_{21} = 0$. At the Cosmic Dawn stars begin to form and emit Lyman – $\alpha$ photons ($z \lesssim 30$). First, the Wouthuysen-Field effect couples the spin temperature back to the gas, while Lyman – $\alpha$ photons are not yet intense enough to heat up the gas above the CMB and therefore $T_s \approx T_K < T_\gamma$. Accordingly a second absorption is expected in the brightness temperature at around $z \sim 20$. Later on, radiation heats up the gas to a point where $T_k > T_\gamma$, and the brightness temperature could become positive for the first time. As the Lyman – $\alpha$ flux increases and reionizes the gas entirely, no hydrogen atoms are left and the brightness temperature should again vanish. The evolution of the brightness temperature is also shown in Fig. 6 where we show in the black line an example model (taken from [1016]) that exhibits all of the main features discussed above. We also indicate by a gray band the large uncertainties in determining the brightness temperature once star formation starts [1027].

One important outcome from the above discussion is that before Lyman – $\alpha$ photons heat the gas, the gas is colder than both the CMB and the spin temperature,

$$T_s \geq T_K. \quad (28)$$

This observation will be important when we later discuss the EDGES anomaly and implication of BSM physics on 21-cm cosmology.

4.2.2. Observation and foregrounds

Before listing some of the important experimental efforts to measure the global (sky averaged) 21-cm brightness temperature, it is important to describe the observational challenges such experiments are facing. The obstacles are, as always, the faintness of the signal relative to foregrounds. A key number that one should keep in mind is that foregrounds are around 3 to 4 orders of magnitude larger than the $\Lambda$CDM expected signal [1019, 1020]. The foregrounds can be generally subdivide into two sources; diffuse galactic radiation, effecting mainly angular scales larger than a degree, and extragalactic radio noise effecting smaller scales. The galactic radiation is generated mainly through synchrotron emission (with a small bremsstrahlung contribution at higher energies) by cosmic-ray electrons traveling in
the Milky Way magnetic field. The power-law spectrum of cosmic ray electrons \cite{227} translates into a (different) power-law intensity of photons, ranging from a few tens of MHz photons to a few tens of GHz. Since cosmic ray electrons are most prevalent in the galactic plane, the intensity of synchrotron photons depends on the direction and is suppressed at higher Galactic latitudes. The extragalactic radio foreground is dominated by active galactic nuclei (AGNs) and star-forming galaxies (SFGs). Accretion of matter onto super-massive black holes in the center of galaxies, in what is known as radio-laud AGNs, releases a copious amount of synchrotron radio emission. Radio emission from SFGs is very similar in nature to the galactic foreground and is also produced via synchrotron radiation from supernovae related high energy electrons. Foreground mitigation techniques usually rely on a spectral difference between the signal and foregrounds. An extensive and detailed account of foregrounds sourced and mitigation is beyond the scope of this white paper and can be found for example in \cite{1019, 1020, 1028}.

\textit{a. Global Signal Experiments and the EDGES Anomaly}  Global signal experiments aim to measure the sky-averaged 21-cm absorption line and therefore do not require a fine angular resolution. For that reason, in contrast to experiments focused on anisotropies, global signal experiments are designed as single antennas. Because of their low spatial and frequency resolution, foreground mitigation based on spectral information is quite challenging and a phenomenological foreground modeling is required. Such modeling is the source of large uncertainties when analyzing the results of such global signal experiments. Below we present a partial list of some of the ongoing attempts to measure the global signal. A more detailed account can be found, for example, in \cite{1029}.

- \textbf{The Experiment to Detect the Global EoR Signature (EDGES)} \cite{1030} is a single dipole antenna, located in Western Australia. As of today, EDGES is the only collaboration to claim to have detected a cosmological 21-cm signal, \cite{1031}. Their 2018 measurement in the 50-100 MHz range showcased an anomalously strong absorption signal centered around 78 MHz (or $z \sim 17$). Their findings led to an extensive body of work aiming to explain their observation, or to scrutinize it. Their finding and its implication for BSM physics will be discussed with some details later in this white paper.

- \textbf{The Shaped Antenna to measure the background RAdio Spectrum (SARAS)} \cite{1032}
FIG. 7. Current experimental status of the global 21-cm brightness temperature. The green region represents the SARAS 3 2σ constraint on the signal [1038] (digitized manually since the data is not yet publicly available). The red error-bars are the LEDA 2σ constraints [1041], while the blue band is the best fit model to EDGES data [1031], taken from [1020]. The ΛCDM prediction band is shown in shaded gray (taken from [1027], same as Fig. 6). The orange line shows the crude lower bound on the brightness temperature defined in Eq. (29) (taken from [1020]).

is a cone shape antenna. Each of its three generations, SARAS 1-3, was located in a different part of India. Measurements by SARAS 2 [1033] in the frequency range 110-200 MHz were used to mildly constrain different astrophysical and cosmological models [1034–1036]. The third generation, SARAS 3, observed in the frequency range 43.75–87.5 MHz during the first quarter of 2020 [1037, 1038]. The result of SARAS 3 is claimed to contradict the EDGES result by more than 95% confidence level [1038]. Results from SARAS 3 are expected to improve the constraints on astrophysical parameters.

• The Large-Aperture Experiment to Detect the Dark Ages (LEDA) [1039, 1040] is located in California, USA. Unlike the previous two examples, LEDA is an interferometer consists of about 250 dipole antennas spread over an area of roughly 200 meters. LEDA is designed to observe in the frequency range 30 to 88 MHz ($z \sim 16 – 34$). LEDA has already been used in 2016 to set constraints on the global 21-cm signal [1041].
constraints on the global signal from LEDA [1041] and SARAS 3 [1038], together with the best fit model and the 99% confidence interval based on EDGES measurement [1031]. This figure makes the tension between SARAS 3 and EDGES evident, and helps in the intuitive understating of Ref. [1038], which claimed to exclude EDGES by over 95% confidence.

The measurement claimed by the EDGES collaboration in 2018 [1031] showcased a surprisingly large absorption signal at $z \sim 17$, and was claimed to be in 2$\sigma$ tension with the $\Lambda$CDM prediction [1042]. The $\Lambda$CDM prediction is, however, subject to a lot of uncertainties, originating mainly from 7 unknown astrophysical parameters (see, e.g., [1027, 1043]). To understand how a 2$\sigma$ deviation can be claimed, regardless of those unknowns, we return to Eq. (28) which, together with the definition of the brightness temperature (26), can be translated to a lower bound on the brightness temperature

$$T_{21} \geq T_{21,\text{min}}(z) \equiv \frac{T_K(z) - T_\gamma(z)}{1 + z} \left(1 - e^{-\tau(z)}\right)\bigg|_{T_s = T_k}. \quad (29)$$

Calculating the above within $\Lambda$CDM and neglecting any heat source for the baryonic fluid results in a robust lower bound on the brightness temperature. The EDGES measurement was anomalously lower than this very aggressive lower limit. Fig. 7 also shows in gray the $\Lambda$CDM prediction band (same band as in Fig. 6). The distance between the gray and blue bands demonstrates clearly the deviation from $\Lambda$CDM.

Immediately after EDGES published their results it attracted a lot of attention both form those wishing to explain it, as will be elaborated shortly, and those who questioned the validity of their analysis. The concerns about the analysis focused on issues related to unknown instrumental systematic, contamination, foreground subtraction and statistical modeling [1044–1048].

Taking the EDGES result at face value, however, calls for an explanation. From Eq. (29) we see that the tension is relaxed if the gas temperature is lowered, or the CMB temperature (in the relevant frequency range) is increased. In the following section we explore BSM solutions to the EDGES anomaly. However, it is worth mentioning some astrophysical processes that were suggested as a possible resolution. Ref. [1049] suggested that the EDGES anomaly could be resolved by positing an excess radio background. However, it was later claimed that this solution fails to explain EDGES [1050, 1051]. Somewhat optimistic models for the formation and evolution of super-massive black holes were also suggested as a possible source of radio background large enough to explain the EDGES anomaly [1052, 1053].
4.2.3. BSM physics and 21-cm cosmology

For BSM physics to leave an imprint on the 21-cm brightness temperature it should somehow affect the evolution of the spin temperature. In what follows we summarize a couple of scenarios that achieve this. Most examples will be given in the context of the EDGES anomaly. However, we will try to emphasize features that would remain relevant even if the EDGES measurement is falsified.

a. *DM - baryon interaction*  
DM-baryon scattering was one of the first BSM implications of 21-cm observation to have been discussed in the literature [1054]. Ref. [1055] addresses the heating of baryons by the drag exerted by the DM fluid on baryons, due to DM-baryon relative bulk motion [1056]. The authors however do note that more often than not, the CDM acts as a reservoir that either immediately cools the baryons (as was originally suggested in [1054]), or heats up the baryons for a very short period before again cooling them down through this ‘dark reservoir’. Importantly, the authors of [1054, 1055] noted that to allow any effect during the Dark Age or Cosmic Dawn, without being in tension with CMB observations [1057–1063], the DM-baryon interaction rate must be enhanced at low velocities (since the gas temperature and relative bulk velocity at CMB is larger than during the Dark Age and Cosmic dawn). After EDGES published their anomalous results, the same idea was used as a mean for cooling down the baryons, thereby lowering $T_{21,\text{min}}$ (cf. Eq. (29)) enough to agree with EDGES [1042]. There the author assumed a Rutherford like cross section between DM and baryons, perhaps originating from DM carrying a small electric charge, i.e., millicharged DM (mDM). It was immediately noticed that the strength of interactions required to reconcile EDGES is in tension with CMB observations [1064, 1066]. It was further established in [1065] that constraints from 5th force experiments [1067, 1068], together with stellar cooling bounds [1069, 1070], lead to the conclusion that the EDGES anomaly could only be addressed if the DM-baryon interaction strength is proportional to the electric charge. It was suggested in [1064, 1066] to consider only a small fraction of mDM such that their abundance is low enough to evade CMB constraints [1062, 1063, 1071], while large enough to sufficiently cool down the gas and reconcile EDGES. The mDM parameter space however is still severely constrained by measurements of the number of relativistic degrees of freedom at BBN [1072], direct detection [1073, 1077], colliders [1078], and supernova 1987A cooling [1079]. The viable parameter space for the mDM explanation of EDGES then
gradually shrank as CMB \[1062, 1063\], BBN \[1080\] and direct detection \[1081\] constraints improved. Recently a cosmologically viable model based on the Stueckelberg mechanism was put forward \[1082\], offering a thermal production mechanism for the mDM that relaxes the EDGES tension, while also mildly relaxing the BBN constraint, thereby allowing for a larger viable parameter space. Ref. \[1083\] opened a whole new parameter space for the mDM explanation of EDGES, by allowing the mDM fraction to interact via a long range force with the CDM bath.

Another scenario, very different in detail but similar in that it used DM to cool down baryons, was presented in \[1084, 1085\]. There, based on \[1086\], it was claimed that if axion DM forms a Bose-Einstein condensate, it should cool down the gas via gravitational cooling.

If EDGES is falsified, as claimed by SARAS \[1038\], measurements of the global signal can still constrain over-cooling generated by DM-baryon interactions \[1087\]. The main challenge would be the large uncertainty on the SM prediction discussed before; for DM-baryon interactions to significantly cool baryons, the interaction rate should exceed the unknown stellar heating rate. In the absence of prior knowledge, one is therefore forced to adopt the strongest conceivable heating rate, thereby diminishing the constraining power of such observation. It is likely that a constraint would be made possible only once the ΛCDM prediction is independently inferred or constrained.

b. Modifications to the Rayleigh–Jeans Tail

As discussed earlier, another way of relaxing the EDGES tension is by increasing the CMB intensity at the low energy tail of the spectrum (the Rayleigh–Jeans tail), relative to which the brightness temperature is defined. An elegant BSM solution to the EDGES anomaly was then suggested based on the conversion of hidden photons \[1088, 1089\] or axions \[1090\] into Rayleigh–Jeans photons. In both cases a resonant conversion of a hidden radiation field (hidden photons or axions) into soft photons amplifies the brightness temperature in accordance with the EDGES observation. We note that excess above the CMB black body spectrum at the Rayleigh–Jeans tail was observed also by ARCADE2 \[1091\], though in much higher energy than what EDGES measured. A hidden to visible photon conversion seems to also give partial explanation to the ARCADE2 excess \[1092\]. The conversion of hidden photons into photons was later studied in greater detail in \[1093–1095\], where the authors considered the effect of inhomogeneities in the electron density which translate into inhomogeneous transition probability from hidden to visible photons. It was also shown that hidden to visible photon conversions could lead
to a distinct spectral feature in the observed brightness temperature \cite{1092}. While those works were presented in the context of EDGES, most arguments could be flipped and lead to constraints on hidden to visible radiation conversion in the relevant parameter region.

c. Direct Spin flips  Another possibility for BSM models to effect the brightness temperature is by directly effecting the spin temperature, namely, the relative abundance of singlet and triplet hydrogen states. For a new spin-flip rate to exist, it seems reasonable to expect that axial interactions with the hydrogen atom are required. The inclusion of a new spin flipping process manifests itself through the addition of a new term in Eq. (25). Direct axion induced spin-flips were the subject of Ref. \cite{1096}. They find that axion-induced spin-flip transitions always increase the spin temperature. It was also realized, however, that even if the axion mass is tuned to match the hyperfine splitting energy gap, the spin-flip rate is very suppressed compared to the CMB induced rate. In \cite{1097}, however, it was shown that in the case that the axion relic density forms a coherent state, induced emissions and absorption of axions can significantly lower the spin temperature. In that case the mass of the axion is not tuned to the hyperfine splitting. However, the axion energy spectrum should have a significant support around the hyperfine energy splitting. An entirely different example for a DM-induced spin flip was introduced in \cite{1098}. There, a fraction of DM density interacts with electrons via an axial vector, allowing for cooling of the gas (much like the mDM case discussed earlier), but also for a direct coupling of the spin to a new effective temperature. Due to the axial nature of the interactions, in a large portion of the parameter space the spin-flip rate dominates, offering a very distinct redshift evolution of the spin temperature.

d. Energy Injections  The last example we wish to discuss is the effect of energy injection, either from decays or annihilation, on the brightness temperature. This is perhaps the most natural BSM implication of a brightness temperature measurement, as it resembles constraints on energy injections derived for example using CMB measurements \cite{1058, 1099–1102}. The basic idea is that energy injected to the SM would heat up the gas, thereby increasing the brightness temperature whenever the spin temperature follows the kinetic temperature. There are two possible ways for energy injection to heat up the gas \cite{1103, 1104}. First, energy injection could increase the ionized fraction, allowing for more efficient Compton scattering, henceforth baryons decouple later from the CMB photons. Later kinetic decoupling means the gas had less time to adiabatically cool, and therefore baryons are
hotter than the ΛCDM prediction. The second option is that energy is injected into the plasma in the form of stable SM particles that initiate a shower that directly heats up the gas. Generically the second option is dominant. The brightness temperature could therefore be used to constrain energy injections coming from, for example, decays [1102, 1105–1109], annihilation [1103, 1106, 1109, 1110] and PBH evaporation [1109, 1111, 1112].

e. Small-scale suppression As a last example we mention how a small-scale suppression to the linear matter power spectrum should effect the brightness temperature profile. The idea is that for the brightness temperature to exhibit an absorption, the spin temperature should couple to the gas. As discussed earlier, at the Cosmic Dawn such coupling is achieved through the Wouthuysen-Field effect, i.e., that the presence of Lyman – α radiation couples the spin temperature to the gas. Therefore, if a Cosmic Dawn absorption is measured at frequencies corresponding with a certain redshift \( z_{CD} \), we conclude that stars should have already produced a significant amount of radiation prior to \( z_{CD} \). Since halo and star formation are halted in models where the matter power spectrum is suppressed at small scales, observational determination of \( z_{CD} \) could be translated into a constraint on such models. Following similar logic, EDGES findings were used to set constraints on Fuzzy DM and warm DM [1113–1115].

4.3. The cosmological lithium problem

[Contributor(s): Josef Pradler]

The epoch of BBN is a crucial moment in the early thermal history of the Universe [1116]. Its outcome, the abundances of light elements, sets the chemical initial conditions for the later periods of recombination and formation of the first stars. Taking the CMB-determined value of the baryon-to-photon ratio \( \eta_{CMB} = (6.104 \pm 0.058) \times 10^{-10} [579, 1117] \) as input, and assuming an uneventful standard cosmic history between redshifts \( z_{BBN} \sim 10^9 \) and \( z_{CMB} \sim 1100 \), makes BBN a parameter-free theory. The predictions of light element abundances are a 25% helium mass fraction \( Y_p \) and \( \mathcal{O}(10^{-5}) \) abundances \( \text{D}/\text{H}, \text{^3He}/\text{H} \) and \( \text{T}/\text{H} \) with the latter beta-decaying into \( \text{^3He} \). With most neutrons locked up in \( \text{^4He} \) and because of the absence of mass-5 and 8 stable isotopes, only \( \mathcal{O}(10^{-10}) \) traces of mass-7 elements, \( \text{^7Li}/\text{H} \) and \( \text{^7Be}/\text{H} \), a tiny fraction \( \mathcal{O}(10^{-14}) \) of \( \text{^6Li}/\text{H} \), and negligible amounts of heavier isotopes are produced; at a later stage, \( \text{^7Be} \) decays to \( \text{^7Li} \) through electron capture.

The CMB+BBN predictions are then to be compared with the observationally inferred
primordial abundances; for reviews see [401, 1117–1122]. Recent qualitative advances may be considered the detection of primordial helium in the CMB alone [1123, 1124] (albeit with a large error-bar), and a much-reduced error to 1% in the determination of D/H = (2.527 ± 0.030) × 10^{-5} in the Lyα absorption system of a high-redshift quasar [1125]. The inference of primordial ^3He/H is problematic and obscured by galactic chemical evolution [1126], but ^3He/D < 1 may be considered as a robust marker because of the monotonicity of this ratio in the low-redshift Universe [1127]. In summary, an impressive overall agreement with the BBN predictions is found, making it a cornerstone of modern cosmology.

The minute traces of ^7Li/H are observed through the 670.8 nm absorption line of the neutral lithium fraction in the atmospheres of old, metal-poor pop-II stars in the galactic halo [1128, 1129]. The flattening, little scatter, and independence of inferred lithium from metallicity suggests the primordial origin at the level ^7Li/H ≃ (1.6±0.3)×10^{-10} [1130, 1131], called the “Spite-plateau” [1132]. Famously, the CMB+BBN prediction is a factor of a few larger, ^7Li/H = (4.72 ± 0.7) × 10^{-10} [1117], putting it at variance with observations at high nominal significance. This long-standing discrepancy is called the “cosmological lithium problem” and has sparked significant interest and led to many speculations about its origin and resolution; see [1133].

Astrophysical solutions and recent observations: lithium that gets exposed to the hotter interiors in stars is consumed. The subsequent reduction of the surface lithium abundance is seen in low-temperature stars that feature large convective zones [1134]. However, standard stellar models predict little reduction in hotter, metal poor stars, but diffusion, rotationally induced mixing or pre-main-sequence depletion may factor in [1135, 1136]. A vital hint towards the role that astrophysics might play is the recent observed scatter of ^7Li/H in the stars at the very metal-poor end of the Spite plateau, resulting in a general drop off from the plateau value [1129, 1131, 1137–1139]. The non-detection of the much more fragile isotope ⁶Li/H [1140, 1141] keeps the stellar depletion option alive, albeit poorly understood and generally requires either an ad hoc mechanism and/or fine-tuning of the initial stellar parameters [1136, 1142]. A conclusive picture of the role of stars in the lithium puzzle has yet to emerge.

Primordial solutions: our understanding of primordial lithium production may be incomplete. The “nuclear option,” i.e., overlooked (resonances in) key reactions [1143, 1144], has now been solidly excluded [1145–1147], closing the window on apparent SM solutions. In

84
turn, interference of new physical processes and/or a departure from the standard cosmic history during the main stages of nucleosynthesis remains a perfect possibility; for a general overview of BBN as a probe for new physics, see [1118, 1148, 1149] and references therein.

At the CMB inferred value of the baryon asymmetry $\eta_{\text{CMB}}$, lithium is predominantly produced in the form of $^7\text{Be}$ through the reaction $^3\text{He}(\alpha, \gamma)^7\text{Be}$ at a sub-Hubble rate. Prospective solutions to the lithium problem require a corresponding reduction of this isotope—without spoiling the overall success of BBN predictions. With the advent of high-precision determinations of D/H [1125], this has become a serious challenge. For example, increasing the rate of the Hubble expansion through the presence of additional radiation degrees of freedom lowers the “freeze-in” yield of $^7\text{Be}$, but only at the unacceptable cost of not only increasing $Y_p$ but also D/H beyond its observational bounds [1149]. A more promising option is instead a change of timing that is brought along by shifting the deuterium bottleneck through a modification of the deuteron binding energy [1150, 1152].

There is a long history of using BBN to probe non-equilibrium processes induced by the decay, evaporation, or annihilation of massive relics [1153–1161] with more recent improvements in [1162–1173]. Although the destruction of $^7\text{Be}$ and $^7\text{Li}$ in the resulting electromagnetic and hadronic showers can yield a net-reduction in mass-7 elements, the spallation of the far more abundant D and $^4\text{He}$ targets is typically catastrophic. Finally, if electromagnetic charged relics $X^-$ are present in an abundance that is similar to that of baryons, bound state effects can even catalyze [1174] nuclear reactions and lead to new lithium depleting channels [1173–1177]. Generally speaking, any such solutions to the lithium problem require rather specific, if not to say, fine-tuned departures from the standard picture. For example, late-time electromagnetic energy injection below the deuteron dissociation threshold of 2.22 MeV but above the $^7\text{Be}$ binding energy of 1.59 MeV against photo-dissociation into $^4\text{He} + ^3\text{He}$ will differentially reduce the mass-7 abundance [1149, 1170, 1178, 1180].

A particular promising avenue to achieve a net reduction of lithium is to increase the neutron concentration to $O(10^{-5})$ after the main stage of BBN ($T \sim 50\text{keV}$) [1156]. It sets in motion the reaction chain $^7\text{Be}(n, p)^7\text{Li}$ followed by the destruction of the created $^7\text{Li}$ isotope with its comparative lower Coulomb barrier, $^7\text{Li}(p, \alpha)^4\text{He}$, resulting in a net reduction of the mass-7 abundance. “Extra neutrons” can, e.g., be created in the decay of some GeV-mass relic $X$ [1169] into pions, muons, or neutrinos that subsequently interconvert $p$ to $n$ through charge exchange such as $\pi^- + p \rightarrow \pi^0 + n$. Again, the problem is that these
“extra neutrons” recombine with protons as well, creating an overabundance of D/H unless such energy injection scenario is carefully adjusted \cite{1169, 1181, 1182}. A more promising approach is to rather “borrow” neutrons at late times by temporarily splitting $D + X \rightarrow n + p$ through the absorption of some relic $X$ with either sufficient mass or energy \cite{1183}. Most released $n$ recombine with $p$ well after the main stage of BBN, leaving the final D/H yield unaltered with comparatively less tuning. The combination of longevity of $X$ plus the required coupling to nucleons make this lithium solution testable at the intensity frontier.

Forty years after discovering the Spite plateau, the cosmological lithium problem stands unresolved. Recent observations reveal a complex picture, with some guaranteed but yet poorly understood role from astrophysics. A new physics origin of the discrepancy remains a viable option, but one that comes with its own challenges.

### 4.4. Small scale galactic structure

[Contributor(s): Hai-Bo Yu]

Galaxies are building blocks of the universe and observations of their structure and clustering provide important information on early-universe cosmology and the nature of DM. For instance, in the 1970s, the neutrinos were thought to be a promising DM candidate within the SM if their mass was $\sim 10$ eV. When the neutrinos decoupled from the SM thermal bath at temperature 1 MeV, they were still “hot,” i.e., relativistic. And their free-streaming scale was so large that density fluctuations below supercluster scales would be erased. In the early 1980s, numerical simulations showed that the structure formation scenario expected in the neutrino-dominated universe, where superclusters form first and they subsequently fragment to produce galaxies, is incompatible with galaxy clustering constraints \cite{1184}.

Substructure clustering of the Milky Way and the Local Group has been used to test the CDM paradigm. Since CDM halos grow via hierarchical mergers of smaller halos, they are rich with substructures. For a Milky Way-like galaxy, CDM simulations predict 100–1000 subhalos that could host galaxies. However, the number of observed satellite galaxies of the Milky Way is much less than that of predicted subhalos. This was originally coined as the Missing Satellites problem \cite{1185, 1186}. More recent studies show that the tension can be largely alleviated or resolved within CDM, see e.g., \cite{1187}. Star formation in subhalos could be suppressed due to environmental effects, and the presence of the stellar disk in the Milky Way could significantly reduce the number of subhalos due to tidal stripping. Recent galaxy
surveys discovered more satellite galaxies associated with the Milky Way. Taking these into account, CDM appears to be consistent with the abundance constraint of the Milky Way satellites.

Turning this around, we can use the abundance constraint to study DM models that have a suppressed matter power spectrum on small scales compared to CDM. For example, Ref. [1188] uses the latest survey data and shows that the lower limit on the mass of thermal relic warm DM is 6.5 keV with the free-streaming length less than $10 h^{-1}$ kpc. For fuzzy DM, the mass must be larger than than $2.9 \times 10^{-21}$ eV and the corresponding de Broglie wavelength is less than 0.5 kpc. Unlike terrestrial searches for DM, those constraints are independent of whether or not DM couples to normal matter.

The matter power spectrum can also be damped by interactions between DM and DR in the early universe [1189, 1190]. The damping scale is set by the inverse of the horizon size when kinetic decoupling occurs, and we can estimate it as $k \sim 10(T_{kd}/ \text{keV}) \text{Mpc}^{-1}$, where $T_{kd}$ is the decoupling temperature. Suppose the scattering cross section is $\langle \sigma_{\chi r} v \rangle$, and write the momentum transfer rate as

$$\Gamma = n_r \langle \sigma_{\chi r} v \rangle \frac{T}{m_\chi}, \quad (30)$$

where $n_r \sim T^3$ is the radiation number density, $T$ the temperature of the thermal bath, and $m_\chi$ mass of DM particles. We obtain $T_{kd}$ by setting $\Gamma = H$, where $H \sim T^2/M_{pl}$ is the Hubble expansion rate and $M_{pl}$ is the Planck mass. Taking $\langle \sigma_{\chi r} \rangle \sim T^2/m_\phi^4$, where $m_\phi$ represents the mediator mass scale for DM-radiation interactions, we have

$$T_{kd} \sim 10 \text{ MeV} \left( \frac{m_\phi}{100 \text{ GeV}} \right) \left( \frac{m_\chi}{100 \text{ GeV}} \right)^{\frac{1}{4}}. \quad (31)$$

For WIMPs, both $m_\chi$ and $m_\phi$ are at the weak scale $\sim 100$ GeV. Thus the damping scale is $k \sim 10^5 \text{ Mpc}^{-1}$ [1191, 1192], which is far beyond current observational limits. However, over the past decade, many novel DM models based on the idea of dark sectors have been proposed [18, 153, 313, 1193, 1194]. These models often predict the existence of a light mediator. In this case, kinetic decoupling from the thermal bath could be delayed, resulting in low $T_{kd}$ [295, 313, 318, 1195, 1197]. For $m_\phi \sim 10$ MeV, $T_{kd} \sim 1 \text{ keV}$ even for a weak-scale DM mass. Thus the relevant damping scale becomes $k \sim 10 \text{ Mpc}^{-1}$, which can be probed with current and upcoming observations. Fig. 8 shows damped linear matter power spectra due to interactions between DM and DR in the early universe (left panel) and their impacts on the number of subhalos in a Milky Way-sized halo (right panel).
FIG. 8. *Left:* Linear matter power spectra for CDM, warm DM and “acoustic damping” DM models. The oscillatory features of the damped DM power spectra are due to interactions between DM and DR in the early universe, analogues to BAOs in the visible sector. *Right:* The corresponding cumulative number of subhalos in a Milky Way-sized halo vs. their maximal circular velocity. For comparison, the black dot denotes the observed number with (dashed line) and without (solid line) coverage corrections. Taken from the ETHOS project [1198].

FIG. 9. Profiles of DM velocity dispersion (left), densities (middle) and shapes (right) predicted in CDM (black) and self-interacting DM (blue). Taken from [1199]. The simulation data are compiled from [355].

Interactions in the dark sector could change the inner structure of DM halos, leading to observational signatures on galactic scales. For example, if DM has strong self-interactions, similar to the nuclear interactions, DM particles can collide and thermalize the inner halo over cosmological timescales. See [1199] for a review on self-interacting DM (SIDM). Fig. 9
shows profiles of DM velocity dispersion, densities and shapes predicted in CDM and SIDM from N-body simulations of halo formation [355]. For a DM halo, the self-interactions produce a shallower density profile and a rounder shape, compared to CDM. In fact, observations of many dwarf and low-surface brightness galaxies favor a shallow density profile in the inner region, in contrast to the cuspy one predicted in CDM [448]. The idea of SIDM was originally proposed to address this core vs cusp problem and the missing satellites problem [353].

Recent studies have shown that SIDM can provide a unified framework to explain diverse DM distributions over a wide range of galactic systems from ultra-diffuse galaxies [1200, 1201], satellite galaxies of the Milky Way [1202, 1203], and spiral galaxies [1204–1206], while being consistent with strong constraints from galaxy clusters [1199, 1207]. As shown in Fig. 10 (left panel), the cross section per mass is necessary velocity-dependent, from a few cm$^2$/g or higher to 0.1 cm$^2$/g towards cluster scales. Interestingly, the inference of the self-scattering cross section from dwarf to cluster clusters can be used to constrain the mass of DM and mediator particles, as illustrated in Fig. 10 (right panel). Evidence for self-interactions point toward a new dark mediator particle that is much lighter than the weak scale [23, 1208–1211]. In addition, the velocity dependence of the cross section could be tested with observations of galactic substructures [1212–1215]. Thus galactic halos act like particle colliders in probing the particle physics nature of DM.

Another unique prediction of DM self-interactions is that a halo could enter a stage of evolution called gravothermal collapse [1216]. As a self-gravitating system, the halo has negative heat capacity and the collisions could transport the heat from the inner to outer regions, resulting in an extremely high central density. For a typical halo, the timescale to reach this collapse stage is much longer than the age of the universe. However, several effects could speed up this process, such as the presence of dissipative self-interactions [1217, 1218], tidal interactions [356, 357, 1219, 1220] and deep baryon gravitational potential [1221, 1223], resulting in striking predictions in galactic systems. Intriguingly, studies have shown that a collapsed SIDM halo could provide a seed for supermassive black holes [350, 1224–1227].

Observations of galactic systems provide a profound probe of the nature of DM and early-universe cosmology in general. This is independent of the assumption that DM has sizable interactions with normal matter. To further advance this research area, we need a collaborative effort among theorists, simulators and observers.
FIG. 10. DM halos as particle colliders. Left: Self-interacting cross sections measured from astrophysical data, including dwarf galaxies (red), low-surface brightness galaxies (blue) and galaxy clusters (green), as well as halos from N-body simulations with the cross section per mass $1 \text{cm}^2/\text{g}$ (gray). The dashed curve is the velocity-dependent cross section from the best-fit dark photon model shown in the right panel. Right: Parameter space for a dark photon model of DM self-interactions, preferred by dwarfs (red), low surface brightness galaxies (blue), and clusters (green). The solid contours denote the combined regions. Taken from [1199].

5. NEUTRINO MASS MODELS

Contributor(s): Marco Drewes

Neutrino masses represent one of the very few pieces of empirical evidence for the existence of physics beyond the SM, and the only firmly established evidence that has been seen in laboratories on the earth. Their explanation certainly requires the existence of new degrees of freedom, generally predicting the existence of new particles. These new degrees of freedom can affect the physics of the universe in various ways, depending on their masses and interactions, including

1) Baryogenesis/leptogenesis: The new particles may also be responsible for the overall matter-antimatter asymmetry in the observable universe [402]. The probably most studied scenario of this kind is thermal leptogenesis [584], but by now many implementations of this idea exist, cf. Sec. 3.1.

2) DM: One or several of the new particles can be DM candidates. This could be a heavy sterile neutrino mass eigenstate itself [31] (cf. Sec. 2.7) or additional particles that are involved in the neutrino mass generation. A review of models implementing the former possibility can be found in [376]. Generally radiative neutrino mass models offer more
freedom to accommodate the latter possibility. The most prominent one of them is probably the scotogenic model [1228], but in the meantime many other scenarios have been studied [1229].

3) **Phase transitions**: If the new particles include an extended scalar sector that spontaneously breaks some symmetry, as e.g. in left-right symmetric models [411, 413, 1230], then this can give rise to a phase transition in the early universe. If this phase transition is of first order, it can potentially lead to gravitational wave signatures [646, 1231–1234].

4) **Hubble tension**: The current disagreement between local and cosmological measurements of the Hubble constant [831] discussed in Sec. 4.1 could be explained by sufficiently light new particles involved in the generation of neutrino masses, in particular sterile neutrinos [1235], Majorons [881] (potentially opening a connection to leptogenesis [883]) and self-interacting neutrinos [1236].

5) **Dark Energy**: In theories in which the accelerated expansion of the universe is driven by a dynamical field, in particular in quintessence type of scenarios [911], it has been pointed out that the origin of neutrino masses may be related to [1237, 1238] this field as well.

6) **Theory of Gravity**: It has been proposed that the origin of neutrino mass may be related to the gravitational anomaly [1239].

7) **BBN and CMB**: If the masses of the new particles are light enough and/or their lifetimes are long enough, they can affect cosmic history at comparably late times. Depending on their mass and lifetime they would contribute to the DM density, radiation density (dark radiation), or impact the primordial plasma through their decays (e.g. by releasing entropy, or by dissociating nuclei or atoms). This can impact on several cosmological observables, including the light element abundances in the intergalactic medium [1165, 1171, 1180, 1240, 1245] and the CMB [1101, 1246, 1249].

The precise way in which the new particles affect the early universe depends on the specific model under consideration. The facts that neither new particles nor clear deviations from
unitarity in the light neutrino mixing matrix have been seen\(^{11}\). suggests that the mass scale \(\Lambda\) of the new particles is much larger than the typical energy \(E\) of neutrino oscillation experiments, permitting to describe neutrino oscillations in the framework of EFT in terms of operators \(\mathcal{O}_i^{[n]} = c_i^{[n]} \Lambda^{n-4}\) of mass dimension \(n > 4\) that are suppressed by powers of \(\Lambda^{n-4}\) \(^{[1252]}\). The Wilson coefficients \(c_i^{[n]}\) in this expansion are tensors in flavour space. The only operator of dimension \(n = 5\) \(^{[1253]}\) is
\[
\ell_L Y_1 \nu_R \tilde{\Phi} - Y_1 v_\Delta \ell_L Y_1 \Sigma_L \tilde{\Phi},
\]
with \(Y_1\) the SM Higgs doubled and \(\ell_L\) the lepton doublet. It generates a Majorana mass term. Dirac mass terms can be generated from operators with \(n > 5\) \(^{12}\). This Majorana mass can be generated in three different ways at tree level \(^{[1254]}\), \(\ell_L Y_1 \nu_R \tilde{\Phi}, \ell_L Y_{1i} \sigma_2 \Delta \ell_L\) and \(\ell_L Y_{1i} \Sigma_L \tilde{\Phi}\), known as type-I \(^{[368, 373]}\) type-II \(^{[373, 606, 607, 1255, 1256]}\) and type-III \(^{[609]}\) seesaw. Here \(\nu_R\) a SM singlet fermion (“right handed neutrino”), \(\Delta\) a SU(2)\(_L\) triplet scalar, and \(\Sigma_L\) a fermionic SU(2)\(_L\) triplet, respectively, that couple to the SM via the coupling matrices \(Y_1, Y_{1i}, Y_{1ii}\). After EWSB they generate contributions \(m_{\nu_R} = -v^2 Y_1 M_M^{-1} Y_1^T, m_{\nu_\Delta} = -\sqrt{2} Y_{1i} v_\Delta, m_{\nu_\Sigma} = -\frac{1}{2} v^2 Y_{1ii} M_S^{-1} Y_{1ii}^T\), with \(M_M\) and \(M_S\) Majorana masses for \(\nu_R\) and \(\Sigma_L\), respectively, and \(v_\Delta\) the expectation value of \(\Delta\). Of course, combinations of these scenarios may coexist, as e.g. in left-right symmetric models \(^{[411, 413, 1230]}\). At loop level many more possibilities exist \(^{[1257]}\), with some of the more prominent ones proposed in \(^{[1228, 1258–1262]}\). The rich menu of scenarios for neutrino mass generation offers many possibility of connecting them to the phenomena \(1)-7\). For instance, leptogenesis was first proposed in the type I seesaw \(^{[584]}\), but can also be realised with scalar \(^{[1263]}\) or fermionic \(^{[611, 1264]}\) triplets or combinations thereof \(^{[1265]}\), cf. \(^{[612]}\) for a review. Radiative models offer many ways to include DM candidates \(^{[1229]}\).

One way to classify neutrino mass models is according to the reason why the neutrino masses are so small compared to those of other fermions in the SM (mass puzzle) \(^{[378]}\): a) \(\Lambda\) is large, b) the entries of the matrices \(c_i^{[n]}\) are small, c) there are cancellations between different terms in \(m_{\nu}\).

a) **Seesaw mechanism.** For \(\Lambda\) far above the SM Higgs VEV \(v\) the \(\mathcal{O}_i^{[n]}\) (and therefore the light neutrino masses) are parametrically suppressed by powers of \(v/\Lambda\). The observation that larger \(\Lambda\) lead to smaller \(m_i\) has earned this idea the name **seesaw**

---

\(^{11}\) Recent data from MicroBooNE \(^{[1250, 1251]}\) disfavours earlier claims that suggest the existence of new light extra states.

\(^{12}\) Of course, in principle the dimension four operator \(\ell_L Y_1 \nu_R \tilde{\Phi}\) can generate a Dirac mass in the same way as all fermion masses are generated in the SM, requiring no new particles except new neutrino spin states, but this would require Yukawa couplings \(Y \sim 10^{-12}\) and an explanation for the absence of a Majorana mass term for the right-handed neutrinos \(\nu_R\).
mechanism (“as $\Lambda$ goes up, the $m_i$ go down”).

b) **Small numbers.** Small entries of the matrices $c_i^{[n]}$ can be achieved in several ways without tuning them “by hand”. A popular is the breaking of a symmetry \cite{509}, which may also help to understand the flavour puzzle. Small Wilson coefficients can also occur if the respective operator is generated at loop level \cite{1252,1257}. Other possible explanation involve the gravitational anomaly \cite{1239}, extra dimensions \cite{1266,1267}, or string effects \cite{1268,1269}.

c) **Protecting symmetries.** No large $\Lambda$ or small $c_i^{[n]}$ are needed if the light neutrino masses are protected by an approximate symmetry related to an approximate conservation of lepton number \cite{1270,1271}. Models that can incorporate this idea include the inverse seesaw \cite{1272,1274}, linear seesaw \cite{1275,1276} seesaws, scale invariant models with a dark sector \cite{1277} and the $\nu$MSM \cite{1270}.

Of course, any combination these mechanisms may be realised in nature. An important point is that neutrino oscillation data is not sensitive to the magnitude of $\Lambda$, but only to the combinations $c_i^{[n]}\Lambda^{n-4}$, and in principle $\Lambda \sim \text{eV}$ would be possible \cite{1278} \footnote{A seesaw scale below the pion mass is strongly constrained by its impact on the late universe \cite{1233,1245,1247,1249}, cf. 7) above, unless the new particles can decay efficiently into a dark sector \cite{1250}. See \cite{459} for a summary of constraints.} Theoretical motivations for a low seesaw scale are e.g. summarised in Sec. 5.1 of \cite{378}. The phenomenological implications of different choices of the scale $\Lambda$ for cosmology and particle have been studied in most detail for the type-I seesaw, cf. e.g. \cite{377}. An appealing phenomenological aspect is that the new particles can be found in direct searches in accelerator based experiments \cite{377,378,610,619,1281}, including DUNE \cite{1282,1283} and future colliders \cite{1284,1285}. This can also help addressing cosmological problems. For instance, the viable parameter space for leptogenesis in the type-I seesaw mechanism overlaps with the reach of various collider and fixed target experiments \cite{602,605,1286}, providing a testable scenario of neutrino masses and leptogenesis \cite{613} by exploiting the complementarity between experiments at difference frontiers \cite{601,1287}. An example for a UV complete \cite{1288} and testable model that can relate neutrino masses, baryogenesis and DM is the $\nu$MSM \cite{375,410}. Low scale leptogenesis in other neutrino mass models \cite{1289} has been studied less, but is feasible. An exploitation of the complementarity of different frontiers is e.g. also promising in the context of DM, cf. Sec. 2.7 and gravitational wave signatures from phase-transitions in neutrino
mass models. Achieving full testability is in practice difficult for models beyond minimal ones like the $\nu$MSM because the dimensionality of the parameter space tends to be too large. This problem can at least partially be alleviated in models that aim not only to explain the smallness of neutrino masses, but also the properties of the neutrino mixing matrix in terms of symmetries that reduce the number of free parameters [1290–1293].

Overall, the manifold possibilities and observational as well as experimental opportunities have made the connection between neutrinos and cosmology an important area of study in the next few years [1294–1296].

6. AXIONS

[Contributor(s): Anson Hook]

Axions and ALPs are one of the most well-motivated new particles [506, 507, 510, 511]. Axions are the most elegant solution to one of the SM’s central mysteries, the strong CP problem, and can simultaneously act as DM [512–514]. Aside from solving the problem of DM and the strong CP problem, axions are also ubiquitous in string theory [1297, 1298], adding additional motivation for their presence. See Secs. 2.9 and 3.6 for the overview of axion DM and baryogenesis from axions, respectively.

One of the main motivations for the QCD axion is that it is a natural solution to the strong CP problem, see e.g. Ref. [1299] for a review. The strong CP problem manifests itself as why is the neutron EDM so small. The neutron is an object that is made of a charge 2/3 up quark and two charge -1/3 down quarks and has a size of $10^{-13}$ cm. As such, one would naturally expect it to have an EDM $d_n \approx 10^{-13} e \text{ cm}$. Applying the exact same expectation to the molecule $H_2O$ results in a correct order of magnitude prediction of its EDM. Experiments have been performed to look for the neutron EDM and have come up negative, finding that $d_n < 10^{-26} e \text{ cm}$. This mismatch between expectation and reality is the strong CP problem. More concretely, there is a parameter is the SM Lagrangian called $\theta$. The neutron dipole moment $d_n$ is proportional to $\bar{\theta}$ and we have experimentally measured that $\bar{\theta} < 10^{-10}$ [1300]. The puzzle of why this CP violating parameter is so small when the other CP violating angle is $\mathcal{O}(1)$ is the strong CP problem.

The QCD axion solution to the strong CP problem can be seen in the context of molecules where an analogous problem occurs. $CO_2$ is another molecule with a positive ion and two negatively charged ions. A natural expectation would be a non-zero EDM, however carbon
dioxide has an EDM of exactly zero! The reason for this is that the angle between the two bonds is dynamical. If carbon dioxide is bent, so that the EDM is non-zero, it relaxes into a straight line configuration where the dipole moment is exactly zero. By complete analogy, in the case of the neutron, the angle of the “bonds” between the up quark and the down quarks is proportional to $\bar{\theta}$. If $\bar{\theta}$ were dynamical, the neutron would relax to the straight line configuration where the dipole moment vanishes. The QCD axion solution works along the line of carbon dioxide. By introducing a new field, the QCD axion, which acts like $\bar{\theta}$, the neutron EDM is dynamically relaxed to zero.

A purely theoretical motivation for the QCD axion and ALPs comes in the form of string theory. ALPs are a generic consequence of extra dimensional theories such as string theory, see e.g. Ref. [1301]. An ALP naturally arises as the 5th component, $A_5$, of a higher dimensional gauge theory. 5D gauge invariance acts as a shift symmetry that renders the axion light, while exponentially suppressed Wilson loops around the extra dimension eventually give the ALP a non-zero mass. Through the phenomenon of anomaly inflow or 5D Chern-Simons terms, the ALP could even have the couplings needed to solve the strong CP problem.

Aside from generic extra dimensional expectations, string theory provides additional sources of axions in the form of antisymmetric tensors such as the NS 2-form $B_2$ or the RR fields $C_0, C_1, C_2, C_3$ and $C_4$. When compactified to four dimensions, these fields generically have many zero modes that are light, with some examples giving tens of thousands of light scalars [1302]. The Chern-Simons couplings of these fields required for the Green-Schwartz anomaly cancellation gives rise to the photon coupling used to look for axions and the gluon coupling needed to solve the strong CP problem.

Aside from being well-motivated, the couplings of axions can provide a deep insight into fundamental questions such as the quantization of electric charge [1303]. The coupling of axions to photons is quantized in units of the fundamental unit of electric charge squared. As a result, depending on how the axion is discovered, it may also act as a Millikan experiment. An example of this can be found in the context of axion strings. Axion strings rotate polarized light by a fixed amount equal to $Ze_{\text{min}}^2$, where $Z$ is an integer, $\alpha$ is the fine structure constant and $e_{\text{min}}$ is the quantized unit of electric charge with 1 being the charge of the proton. If we were to observe the quantized rotation of light due to an axion string, it would immediately teach us about the quantization of electric charge.
Even if axions are not present today, their presence in the early universe could provide exciting experimental signatures such as gravitational waves. One of the most commonly explored production mechanism for stochastic gravitational waves is that of a string network, see Ref. [1304] for a review. Axions are angular fields and thus support topological defects around which they rotate by \(2\pi\). In the early universe, strings form a scaling solution that tracks the total energy density of the universe. If \(f_a \gtrsim 10^{14}\) GeV [1305], these strings radiate gravitational waves that would be observable at detectors such as the pulsar timing array, LISA and LIGO. Observation of these gravitational waves would provide insight into the earliest times of our universe. Information that could be gleaned from these observations include important information such as the equation of state of the early universe.

Going back even further in time, axions could also play an important role in inflation. One of the most common approaches to inflation, the model of natural inflation [1306], utilizes an axion playing the role of the inflaton. In low scale models, axions often play the role of the curvaton. These presence of axions during inflation can provide interesting observational signatures. Parametric resonance production of photons from axion photon couplings gives a non-gaussian signal so loud that warm inflation models centered around this coupling are essentially ruled out [1307]. Other non-gaussian signals, such as those from the cosmological collider physics program, are enhanced when axionic curvatons are present [1308]. The presence of axions during inflation can easily lead to non-gaussian signals that are well within reach of next generation experiments.

7. SUPERSYMMETRY

SUSY is a unique extension of space-time symmetry [1309] that relaxes the EW hierarchy problem [1310–1314], provides the LSP as a DM candidate [1313, 1315, 1316], can explain the baryon asymmetry of the universe via the Affleck-Dine mechanism [565, 1317, 1318], and leads to precise gauge coupling unification [1312, 1319, 1323]. It can also stabilize intermediate symmetry breaking scales required in other extensions of the SM, such as the PQ symmetry breaking scale. SUSY is thus a framework that can address several problems of the SM and beyond simultaneously.

At the time of writing of this white paper, SUSY particles have not been observed at collider experiments, which requires that their masses are above the EW scale. As a result,
within the MSSM framework the observed EW scale requires tuning of the parameters of the theory at the level of $10^{-2}$ or more. This is called the little hierarchy problem. Regardless, SUSY can still explain the large hierarchy between the Planck or unification scale and the TeV scale. Given the strong motivations for SUSY listed above, it would be reasonable to continue to pursue its phenomenological and cosmological implications.

In this section, we review cosmological aspects of SUSY theories. Sec. 7.1 reviews neutralino or gravitino DM in the (N)MSSM. Sec. 7.2 discusses implications of possible hidden sectors to DM and baryogenesis in SUSY. Sec. 7.3 outlines Affleck-Dine baryogenesis, and Sec. 7.4 discusses the production of the LSP or Q-ball DM in Affleck-Dine baryogenesis.

7.1. Neutralino/Gravitino DM

[Contributor(s): Nausheen R. Shah]

In R-parity conserving SUSY theories, the LSP is expected to be absolutely stable. The minimal content of a SUSY theory would include superpartners of all the SM particles, and is known as the Minimal Supersymmetric Standard Model (MSSM) (see for eg. Refs. [1324, 1325]). Further, unlike the SM, SUSY requires at least two Higgs doublets, one coupling to all the up-type fermions and the other to down-type fermions. Therefore, at a minimum, SUSY would include fermionic superpartners of these Higgs bosons and the SM gauge bosons. In the presence of non-minimal field content, there may be additional neutral fermionic states. A popular example is the Next-to-Minimal Supersymmetric SM (NMSSM) (see for e.g., Ref. [1326]) which includes an additional singlet superfield coupling only to the Higgs superfields. Any one of the weakly charged, electrically neutral set of these neutralinos can be excellent thermal WIMP DM candidates. A key difference between a SUSY neutralino DM candidate and a similarly charged simplified model DM candidate is the expected presence of a set of correlated states in SUSY.

As has been discussed in Sec. 2.1, the standard thermal freeze-out WIMP paradigm is considered to be highly predictive: The DM annihilation cross-section determines the relic density, $\Omega h^2$, which is further related by crossing symmetry to what one may expect to observe at DM-nucleon scattering direct detection experiments, $\sigma_{DD}$. Additionally, there is

Apart from providing a possible resolution to the so-called $\mu$-problem [1327, 1328], the observed mass and SM-like nature of the Higgs boson at the LHC may be realized easily in the NMSSM [1329, 1330] without the need of decoupled heavy Higgs bosons or large radiative corrections from stops.
a dedicated program of direct searches at the LHC, probing the presence of neutralinos in various decay modes. Then, the absence of signals so far in both direct detection experiments and at the LHC seems to imply a tension within the standard WIMP paradigm for neutralino DM candidates. However, it should be stressed that the crossing symmetry relating the annihilation cross-section responsible for setting the relic density and the direct detection cross-section is really an EFT realization at most. If there are multiple propagating degrees of freedom present at the weak scale (as may be expected in SUSY), then it is not really appropriate to integrate them out, and their presence may significantly change the model predictions. Additionally there may be different degrees of freedom in final or initial states relevant for processes setting the relic density versus the DM-nucleon scattering cross-section and/or signals at the LHC, relieving perceived tensions and negating the naive interpretation of the DM WIMP paradigm in SUSY.

Due to SUSY, the couplings of the neutralinos to SM particles are governed by the corresponding SM values, i.e., gauge couplings for the bino and wino (superpartners of the hypercharge and $SU(2)$ gauge bosons), and the SM Yukawas for the Higgsinos (superpartners of the Higgs bosons). However, the (generally assumed Majorana) bino and wino masses are generated via SUSY breaking, and are therefore free parameters in the low energy theory, as is the SUSY Higgs mass parameter, $\mu$, governing the mass of the Higgsinos. Generally the mass eigenstates of the neutralinos are not pure weak eigenstates, instead

$$\tilde{\chi}_i = N_{i1} \tilde{B} + N_{i2} \tilde{W} + N_{i3} \tilde{H}_d + N_{i4} \tilde{H}_u + N_{i5} \tilde{S},$$

where $i = \{1, 2, 3, 4, (5)\}$, the $N_{ij}$ denote the bino, wino, Higgsino-up, Higgsino-down and singlino (the superpartner of the singlet scalar if in the NMSSM) component respectively of the $i^{th}$ neutralino, and the mass-ordered lightest state, $m_{\chi_1}$, is the DM candidate. The mixing between the bino, wino and the singlino is mediated by the Higgsinos.

Since the couplings of the winos and Higgsinos are large and they are further accompanied by almost mass-degenerate charged states (the charginos), consistent relic density is obtained for DM masses of the order of the TeV scale \cite{1331,1333}. However, weak scale bino or singlino DM candidates may realize an observationally consistent relic density without necessitating TeV scale winos and Higgsinos. An observationally consistent relic density for either bins or singlinos may be obtained by a wide variety of processes \cite{1334,1336}. There may be $s$-channel resonant annihilation if $2m_{\chi_1}$ is approximately equal to the mediator mass \cite{1337}. 


FIG. 11. *Left:* Regions in the $\mu - m_{\chi_1}$ plane that produce a relic abundance $\Omega h^2 = 0.12 \pm 50\%$ for different values of $\tan \beta$. The red, green and blue regions correspond to $\tan \beta = 10$, 20, and 60, respectively (corresponding to the Higgs resonance), while the purple region corresponds to the $Z$ resonance which is approximately independent of $\tan \beta$. The lower gray shaded region is excluded by SD-DD constraints set by LUX, which are again approximately independent of the value of $\tan \beta$ for moderate to large values of $\tan \beta$. *Right:* Contours of the SI-DD scattering cross section $\sigma_{SI}^{DD}$ in the $M_H - \mu$ plane for $\tan \beta = 20$ with fixed $m_{\chi_1} = 61.7$ GeV. The narrow black region is excluded by SI-DD constraints set by XENON1T. Taken from [1341].

Some of the most stringent constraints on WIMP DM originate from the current limits on the spin-independent DM-nucleon scattering cross-section (SI-DD) [1357]. In SUSY, this cross-section is mediated by the CP-even Higgs bosons and the squarks, superpartners of the quarks. It is interesting to note that there are significant regions of parameter space, so-called “blind spots”, where $\sigma_{DD}^{SI}$ is highly suppressed due to either a cancellation in the coupling of the Higgs boson to the neutralinos, or due to destructive interference.
between the different mediators [1336, 1339, 1341, 1350, 1358–1360]. On the other hand, the spin-dependent DM-nucleon scattering cross-section (SD-DD), while currently having significantly weaker experimental limits [1361], does not suffer from either of these suppressions. The $\sigma^{SD}_{DD}$ is mediated only by the Z gauge boson, and the $Z\chi_1\chi_1$ coupling is provided via the Higgsino component of $m_{\chi_1}$. Hence, if nature lives in proximity to the blind spot, discovering a signal of DM via SI-DD might be very challenging, but we should be able to probe this region of parameter space via SD-DD in the near future [1362–1364]. An example of such a case is showed in Fig. 11 taken from Ref. [1341]. As shown in the left pane, an observationally consistent relic density may be obtained via resonant annihilation at the Z or Higgs poles, denoted by the shaded bands for different values of tan $\beta$ shown in the $\mu$ vs. $m_{\chi_1}$ plane. Current SD-DD limits, shown as the grey shaded region labeled LUX, provide a weak constraint on $\mu$. The SI-DD for an exemplary DM mass is portrayed in the right panel. The yellow/red region denotes the blind spot, and the entire plane shows its suppression due to destructive interference between the Higgs and the squark contributions. The black sliver at the bottom right denotes the current SI-DD limits.

If DM can annihilate into SM particles, it stands to reason that colliding SM particles may produce DM, which would appear as missing energy in collider experiments. Such searches have been undertaken extensively at the LHC. However, the LHC is a hadronic machine. As such, the production of weakly coupled states directly is luminosity limited. Current searches, both at ATLAS and CMS, place bounds on neutralinos and charginos mostly of the order of a few hundred GeV, which also depend on assumptions about the presence of additional sleptons or squarks. The strongest bounds are on wino-like neutralinos/chargino pairs, which may extend up to 8-900 GeV for a massless neutralino LSP [1365, 1366]. However, sensitivity is significantly reduced for compressed spectra [1367, 1368], which may well be regarded as the most relevant regime for a host of SUSY DM scenarios where co-annihilation occurs. At the High-Luminosity LHC, it is expected that wino-like neutralinos may be probed up to $\sim 1200$ GeV [1369, 1371].

Another interesting possibility arises when combining supersymmetry with gravity, which necessitates gauging supersymmetry, leading to Super Gravity (SUGRA) [1372]. In this case, an LSP gravitino (superpartner of the graviton) may naturally arise as a DM candidate. However, the mass scale may vary anywhere from eV to multi-TeV depending on the SUSY breaking and mediation scheme. Naively there are several cosmological problems in such
scenarios. For example, because of its extremely feeble interactions, if the gravitino was thermalized in the early universe, the freeze-out of the gravitino abundance occurs in a relativistic regime, and the gravitino mass $m_{3/2}$ should be less than approximately 1 keV to avoid overclosing the Universe \cite{44, 1315, 1373, 1374}. Therefore a low reheating temperature is required to dilute the overabundance, which is problematic for thermal leptogenesis for example \cite{1375}. Other options include models with a split spectrum, where only the SM and the gravitino (and some subset of the neutralinos) is present at the TeV scale \cite{49, 1376}. However, there are generally either additional messenger particles or ordinary SUSY particles with masses larger than the gravitino, which may be produced and be in thermal equilibrium, earlier in the evolution of the Universe \cite{1377}. In this case, gravitino abundance is obtained via the freeze-in mechanism due to the late time decays of the other SUSY particles, often called the “superWIMP” scenario, where the gravitino mass may be at the weak scale \cite{1378–1380}. While there are constraints due to, for example, too much energy released during or after BBN, or too many light degrees of freedom, such scenarios can be very predictive, precisely due to these constraints. Interestingly, models of gravitino freeze-in DM from such late decays might be relevant for reducing the recent Hubble tension \cite{1381}. Direct detection of gravitino DM is challenging. However, gravitinos may be produced in cascade decays at the LHC, and hence may be probed in various channels involving missing energy, including searches for long lived particles \cite{1382, 1384}. While such results are highly model-dependant, they provide an exciting opportunity to explore the SUSY parameter region consistent with gravitino DM.

Lastly it should be noted that gravitino DM consistent with the observed relic density may also be obtained in R-parity violating SUSY models, where due to its Planck-suppressed couplings, gravitinos may naturally have a lifetime longer than the age of the Universe. In such cases, apart from signals at the LHC, there may be potentially interesting signatures in indirect detection and/or suppressed inelastic signal in direct detection \cite{1385}.

7.2. Dark matter / Baryogenesis from hidden sectors

[Contributor(s): Bibhushan Shakya]

While most early-universe studies of SUSY frameworks focus on weak-scale SUSY with R-parity in the context of the MSSM or one of its known extensions, it is also plausible that none of these aspects might be realized in nature: weak-scale SUSY and SUSY WIMP
DM are severely constrained by LHC and indirect/direct-detection measurements, R-parity is not a necessary ingredient for SUSY and may be only approximate [1386], and there are compelling theoretical arguments for the existence of other hidden/secluded sectors [78, 1387], which could also contain DM. It is thus important to examine the implications of such considerations for DM and baryogenesis, which can lead to observational opportunities very distinct from those expected in canonical SUSY frameworks.

**Dark Matter**

If R-parity is not exact, a new symmetry is needed to stabilize DM (an exception being the gravitino LSP, which is naturally long-lived). Such additional symmetries, as well as new DM candidates, can readily be realized in hidden sectors. If DM is to achieve the correct relic density through the standard freeze-out mechanism (the WIMP miracle), it is interesting to ask how the hidden sector knows about the weak scale, which appears to be an accidental scale of the SM rather than an underlying theory of nature. This coincidence of scales can be explained within SUSY in the form of, e.g., gravity-mediated SUSY breaking, which can endow both the visible and dark sectors with common energy scales (see [1388] for more detailed and quantitative discussions).

The simplest constructions involve dark supermultiplets, where either the scalar or fermion component can be DM, stabilized by a hidden $U(1)'$ symmetry (e.g., [1388–1391]). This naturally allows for kinetic mixing between the hidden and dark sectors; in the SUSY theory, this also gives rise to gaugino and Higgs portals, opening interesting phenomenological prospects (see e.g., [1392, 1393]). Furthermore, there naturally exist superpartners close in mass in the hidden sector, hence DM freeze-out as well as annihilation tend to involve multiple interactions with comparable strength: e.g., DM can annihilate not just into dark gauge bosons but also dark gauginos. Dark matter annihilations also tend to feature multiple cascade steps before reaching the standard SM final states, which modifies indirect-detection signals expected from such DM candidates; in general, CTA will be able to probe several realistic configurations in such setups [1391]. Such generic interplay of multiple interactions and states offers several new possibilities for early-universe cosmology as well as present-day indirect detection.

**Baryogenesis**

R-parity violating (RPV) SUSY has been long known to provide viable baryogenesis mechanisms via RPV decays of heavy superpartners in the early universe [735, 736, 1386]
To satisfy the out-of-equilibrium condition necessary for baryogenesis, such particles must either be extremely heavy, yielding no observational prospects, or have suppressed decay widths, which requires unnatural mass spectra (superpartners separated by several orders of magnitude in mass) or extremely small couplings, which, however, additionally also suppresses the branching fraction of such decays into baryon number-violating channels. Decays of hidden sector superpartners offer a natural remedy to this conundrum, as the small portal coupling involved in the decay naturally suppresses the decay width. Crucially, the hidden sector also offers other advantages: in general, baryogenesis with decays of light particles is severely constrained by several observations, e.g., in SUSY frameworks, EDM measurements strongly constrain large CP phases for particles below the PeV scale \[1398, 1399\]. However, if the large CP phase needed for baryogenesis is contained in the hidden sector, EDMs are doubly suppressed by the tiny portal coupling between the two sectors, alleviating the observational constraints.

A specific realization of the above ideas is gaugino portal baryogenesis \[1400\], which makes use of gaugino mixing between a hidden sector gaugino $\tilde{B}'$ and the bino $\tilde{B}$ of the MSSM, which can produce the baryon asymmetry of the universe consistent with all observations for $\tilde{B}'$ masses as low as 10 GeV. Such frameworks can give rise to observable signals at low energy experiments, such as dinucleon decays or neutron-antineutron oscillations \[1401\], as well as at high energy colliders in the form of long-lived hidden sector particles that decay via RPV couplings.

### 7.3. Affleck-Dine baryogenesis

[Contributor(s): Masaki Yamada]

Affleck–Dine baryogenesis (ADBG) is a mechanism that generates baryon asymmetry via the dynamics of a flat direction charged under $B - L$ (or $B$ for the case with formation of long-lived Q-balls) in SUSY models \[565, 1317, 1318\]. Flat directions are linear combinations of scalar fields, such as squarks (the scalar superpartner of quarks), so constructed that F-term and D-term potentials are absent in the renormalizable level \[1402–1405\]. They only have soft SUSY-breaking terms and an F-term potential coming from a higher-order superpotential in vacuum. A typical flat direction that is widely used in the literature is a linear combination of $\tilde{u}_R^i$, $\tilde{d}_R^i$ and $\tilde{d}_R^k$ ($j \neq k$), where the upper indices represent flavors. The list of MSSM flat directions is given in Ref. \[1406\]. The flat direction that is used to
generate baryon asymmetry is called an AD field.

Because the AD field has a very flat potential by construction, it can have a large field value during inflation. The finite vacuum energy during (and even after) inflation breaks SUSY so that a flat direction obtains an effective mass of the order of the Hubble parameter via SUGRA effects and Planck-suppressed operators $[1318,1407]$. If the mass term is tachyonic, the flat direction has a large VEV, which is expected to be stabilized by the F-term from a higher-order superpotential. Here we note that the same superpotential provides an A-term that breaks CP and $B - L$. As the Hubble parameter decreases to the soft mass after inflation, the flat direction begins to oscillate around the origin of the potential. Furthermore, when CP and $B - L$ are broken by the A-term, the scalar field starts to rotate in the phase space at the onset of oscillation and generates $B - L$ asymmetry.

Depending on the potential of the AD field, small perturbations of coherent oscillations may grow to form non-topological solitons called Q-balls $[488,1408-1413]$. The cosmological scenario of this case is explained in Sec. 7.4. If Q-balls do not form, the AD field is dissipated into quarks through scattering with the heat bath, and the $B - L$ charge of the AD field is converted into a $B - L$ charge of the SM particles (see, e.g., Ref. $[1414,1415]$ for the dissipation process). Subsequently, the $B - L$ asymmetry will be converted to $B$ asymmetry via the EW sphaleron process $[406]$. Considering the effect of $B - L$ and/or lepton flavor violating interactions, if they exist and are efficient after the dissipation of the AD field, may also be necessary $[789,790]$.

The detailed dynamics of the AD field, such as the timing of coherent oscillations and ellipticity of the rotation in the phase space, depend on higher-dimensional operators, SUGRA effects $[1416,1419]$, thermal effects $[1420,1422]$, and low-energy potential in vacuum. Thus, there are many possibilities even in the MSSM. In particular, the resulting amount of baryon asymmetry depends on not only the flat direction but also on the mediation mechanism of SUSY breaking. For example, the soft term from the gauge-mediated SUSY-breaking effect is strongly suppressed when the field value of the AD field is larger than the messenger scale $[1409,1423]$. Furthermore, as the A-term is strongly suppressed in the gauge mediation, the efficiency of the ADBG is relatively suppressed in this model. The formation of a non-topological soliton, called Q-ball, also drastically changes the scenario, as explained in Sec. 7.4.

The ADBG in a non-SUSY model may be considered by introducing a $B - L$ charged scalar
field (see, e.g., Ref. [1424] for a NGB as the AD field), in which case the initial condition of the scalar field may be determined using the stochastic dynamics during inflation [1425–1427]. The fluctuation of the initial VEV may result in the primordial fluctuation of the Universe like the curvaton scenario [1428–1432], which requires that the AD field decays after it dominates the energy density of the Universe.

**Smoking-gun signals**

As the AD mechanism uses higher-dimensional operators and is realized at very high temperatures, verifying it directly is challenging. However, the mechanism may be supported using the observations of its smoking-gun signals.

Low-energy SUSY theory places an upper limit on the reheating temperature of the universe to avoid gravitino overproduction [44, 1162, 1165, 1433–1436], which makes implementation of other baryogenesis mechanisms, such as thermal leptogenesis [584], difficult. On the other hand, the ADBG naturally works in SUSY theory without a high reheat temperature; therefore, the discovery of SUSY particles on the TeV scale is a relatively good support for the AD mechanism. However, the AD mechanism is generically a high-scale phenomenon that does not require a low SUSY-breaking scale.

When the AD field has a large field value during inflation, its phase component may have quantum fluctuations. As the resulting amount of baryon asymmetry depends on the initial phase of the AD field, the quantum fluctuations of the phase result in baryonic isocurvature fluctuations [1437–1439]. They are constrained by the observation of the CMB temperature anisotropies [109], which in turn puts constraints on the AD field and inflation models [1440, 1441]. If DM is also produced from the AD field, the DM isocurvature perturbations can be compensated by the baryon isocurvature perturbations [1442].

The AD leptogenesis by $LH_u$ flat direction [1317] is interesting because its higher-dimensional operator is the dimension-5 operator, $(LH_u)^2/\Lambda$, motivated by the neutrino oscillations. In this case, it is known that the thermal effect is important, and the scenario pins down the lightest neutrino mass and predicts the rate of the neutrinoless double beta decay [1422, 1443, 1444].

**Relations to other cosmological problems**

As the ADBG is a powerful mechanism that works in most situations and can produce even very large baryon asymmetries, it can be considered with other cosmological issues.

The ADBG is compatible with some scenarios that require a low reheating temperature
and large baryon asymmetry. For example, low-energy SUSY models require a low reheating temperature to avoid gravitino overproduction \[44, 1162, 1165, 1433, 1436\]. Additionally, a model with an extra dimension may also require a low reheating temperature to not overproduce Kaluza-Klein modes \[1445, 1446\]. The cosmological moduli problem may be avoided by thermal inflation, which dilutes not only the moduli but also the baryon asymmetry to the desired value \[1447, 1449\]. Even when the moduli are unstable and decay early, the ADBG may yield the desired value after dilution by modulus decay \[1450, 1453\].

There have been models that identify the AD field as the inflaton in SUSY \[1454, 1455\] and non-SUSY \[1456, 1461\] models. Although many attempts have also been given to identify the AD field as the curvaton in SUSY models, they confront several problems such as too fast dissipation and too large baryon isocurvature perturbations \[1462, 1467\]. A consistent model for the curvaton was recently proposed in Ref. \[1442\] to naturally compensate isocurvature perturbations by DM. Even if the AD field is not an inflaton nor a curvaton, it may affect the inflaton potential via SUGRA effects and modify the prediction of inflation models \[1418, 1444, 1468\].

Several scenarios that relate ADBG to DM production have been proposed. Refs. \[1469–1471\] have considered the case with a long-lived AD field without Q-ball formation. The case with Q-ball formation is discussed in Sec. \[7.4\] the late-time decay of Q-balls provides a non-thermal production of DM. The AD mechanism can also be used to generate some asymmetry (which may be related to \(B - L\)) in a dark sector \[1472, 1473\]. Those scenarios may explain the coincidence of the energy densities of baryon and DM because they are simultaneously generated from a single origin.

There is also a scenario to create PBHs from large baryon density fluctuations using the AD mechanism \[1474, 1477\]. The AD field is assumed to have a stochastic distribution around the origin during inflation. After inflation, some part of the AD field is trapped at the origin while some at a large field value. Only the latter part of the AD field can generate baryon asymmetry. As a result, highly inhomogeneous baryon asymmetry is generated, which lead to PBH formation. A similar but different scenario is proposed to generate gravitational waves from short-lived topological defects for the AD field \[1478, 1479\]. However, those scenarios cannot create homogeneous baryon asymmetry because the total baryon charge is zero.
7.4. DM production after ADBG

[Contributor(s): Masaki Yamada]

The AD baryogenesis may be followed by the formation of Q-ball, which itself can be a candidate of DM or can non-thermally produce neutralinos and gravitinos via its decay. A Q-ball is a non-topological soliton constructed using a complex scalar field with a conserved global U(1) charge [487, 1408, 1480-1483]. Its stability is guaranteed by the conservation of the U(1) charge. The condition for the existence of this solution is given by

$$\text{Min}_\phi \left[ \frac{V(\phi)}{|\phi|^2} \right] < \frac{\partial^2 V(0)}{\partial \phi \partial \phi^*},$$

where $V(\phi)$ is the scalar potential. This indicates that it is energetically favored for the scalar field to be localized when its potential is shallower than quadratic.

Q-balls form in ADBG if the potential of the AD field satisfies the above condition [488, 1410-1413]. The potential depends on the mediation mechanism of SUSY breaking. For gravity mediation, the soft mass term obtains a logarithmic correction via renormalization group running (see e.g., Ref. [1324]). If the beta function of the soft mass is negative, which is the case if the contributions from Yukawa interactions are smaller than those from gauge interactions, condition (33) is satisfied. For gauge mediation, if the scalar field has a scale larger than that of the messenger, it suppresses the mediating mechanism, resulting in a strong suppression of the magnitude of the soft mass term [1409, 1423]. Therefore, the effective potential of the AD field is flattened at a large field value; in this case, condition (33) is satisfied.

If the potential of the AD field satisfies (33), the scenario of the ADBG is modified as follows. After the onset of oscillation of the AD field, the small perturbations grow to form condensate fragments, which result in Q-balls [488, 1410]. According to numerical simulations, most of the generated baryon charges would be contained in Q-balls [1411-1413, 1484]. To explain the baryon asymmetry of the universe, the baryon charges contained in the Q-balls should be released into the SM plasma by the evaporation [1485, 1487] and/or dissipation [488, 797, 1488] of Q-balls. The energy of the Q-ball per unit charge depends on the potential of the AD field and cannot be much smaller than the soft mass of the AD field in gravity mediation, while it can be as small as the gravitino mass in gauge mediation. If the energy of the Q-ball per unit baryon charge is smaller than the baryon mass, the Q-ball is stable and can release its charge only by dissipation via scattering with the thermal plasma.
Failing that, the Q-ball can evaporate into quarks or baryons from the surface.

**Case with long-lived Q-ball**

Q-ball can decay (or evaporate) into quarks from its surface if it is kinematically allowed. Because the Q-ball contains a large baryon number in a very small region, its evaporation rate may be suppressed by the Pauli exclusion principle of daughter particles. The rate of this evaporation process is calculated in Ref. [1485], and the flux of quarks is indeed saturated by the Pauli exclusion principle on the Q-ball surface. The annihilation of two squarks (that are components of the Q-ball) into two quarks is also efficient near the surface of the Q-ball [1487]. Gravitinos can also be produced. However, its production rate is not saturated by the Pauli exclusion. Therefore, the total Q-ball decay rate is much smaller than that of the AD field and can have a very long lifetime. This provides a late-time entropy production and non-thermal production of DM.

As a Q-ball consists of squarks, its evaporation can produce light SUSY particles as well as quarks. This provides a scenario of co-genesis for baryon and DM, which may be able to explain the coincidence of the energy densities in the Universe. A non-thermal production of wino/higgsino DM from Q-ball evaporation followed by an efficient annihilation of wino/higgsino was considered by Refs. [1489–1491]. Another scenario without dark-matter annihilation is proposed in Ref. [1492–1494]. It is refined in Refs. [1495–1496] by the fact that the ratio of the number density of SUSY particles and quarks is of order $10^{-2}$–$3$ owing to the Pauli exclusion principle on the Q-ball surface and total degrees of freedom of quarks. This indicates that the LSP with a mass of the order of the EW or TeV scale can explain the coincidence of the energy densities of DM and baryon without fine-tuning or artificial assumptions on, for example, the DM mass.

In a gauge-mediated SUSY-breaking model, the gravitino is the LSP and can be DM. The evaporation rate of the Q-ball into gravitinos was calculated by Ref. [1487]. Although the parameter space is not large, both DM and baryon asymmetry can be explained by non-thermal gravitino production from Q-ball evaporation after ADBG [1497–1500].

It has been reported that gravitational waves are emitted during the formation of a Q-ball [503]. However, the energy density of Q-ball has to be so large as to be cosmologically unfeasible in order to emit an observable amplitude of gravitational waves [1502]. It has been also reported that induced gravitational waves from scalar perturbations are efficiently enhanced if Q-balls dominate the Universe. The energy density of Q-balls falls off
at $a^{-3}$, which is slower than that of radiation. Thus, if the energy density is large enough, Q-balls may dominate the universe before their complete decay. The decay of the Q-balls is saturated by the Pauli exclusion principle on the Q-ball surface and hence is larger for a smaller Q-ball. This results in their sudden decay and rapid transition from the early matter-dominated epochs to radiation-dominated epochs. Consequently, the spectrum of induced gravitational waves from scalar perturbations can be significantly enhanced [1503–1506].

**Case with stable Q-ball**

If the Q-ball energy per unit baryon number is smaller than the baryon mass, the Q-ball is stable, in which case only a small fraction of baryon charge can dissipate into the SM particles via the scattering with the thermal plasma on the surface of the Q-ball. This is the case with a relatively large Q-ball in gauge-mediated SUSY-breaking models.

The stable Q-ball can be DM, and its energy density can be consistent with the observed amount of DM [1413, 1507]. However, the dissipation of stable Q-balls does not produce a sufficient baryon number into the thermal plasma to explain the baryon asymmetry of the Universe. Thus, it may be necessary to introduce another AD field that produces baryon asymmetry without forming stable Q-balls [1508].

A stable Q-ball can absorb quarks because it is energetically favored [489, 490]. When a nucleon collides with a Q-ball, it dissociates into quarks, is converted to squarks in the surface layer of the Q-ball, and emits pions of energy of order 1 GeV. This is known as the Kusenko-Kuzmin-Shaposhnikov-Tinyakov (KKST) process. From the phenomenological point of view, this is similar to the Rubakov-Callan effect for magnetic monopoles [1509, 1510], so that the stable Q-ball can be detected by Cherenkov detectors [1511], such as IceCube [1512], Baikal [1513], KM3NeT [1514], and Hyper-K [1515].

The KKST process also leads to astrophysical bounds from the destruction of neutron stars by a Q-ball. Once a Q-ball is captured by a neutron star, it absorbs the neutrons and destructs the neutron star on a time scale much shorter than a billion years [489, 491, 1516]. If there is no upper bound on the Q-ball charge, this excludes Q-balls as being the DM. However, the charge of the Q-ball is saturated at a finite value owing to the presence of the $B$ (or $B-L$) violating A-term [1508, 1517, 1519], which is introduced to generate baryon asymmetry, as explained in Sec. 7.3. Therefore, astrophysical bounds can be avoided if the saturated value is not sufficiently large to destroy neutron stars.

**Case with charged Q-ball**
A Q-ball with a non-zero electric charge is called a charged or gauged Q-ball [1520]. This can be realized if a large Q-ball has both baryon and lepton numbers, such as the case of a $\tilde{u}_R\tilde{u}_R\tilde{d}_R\tilde{e}_R$ flat direction of an AD field. The baryon part of such a Q-ball can be stable, whereas the lepton part can decay into electrons. The asymmetry of the baryon and lepton numbers results in a non-zero electric charge of the Q-ball [1521–1525]. The electric charge is saturated by $\alpha^{-1} \simeq 137$ via the Schwinger effect.

The charged Q-ball can be DM, and it captures electrons if its electric charge is positive, or it captures protons and He$^{+2}$ if its electric charge is negative. The ion-like objects with electric charges of $+\mathcal{O}(1)$ are likely to be relics in the present universe [1524, 1525]. These heavy ion-like objects are investigated in many experiments (see Ref. [1511]). The most stringent constraint comes from the absence of the trail of heavy atom-like or ion-like objects in $9 \times 10^8$-year old ancient mica crystals [1526, 1527].

8. COMPOSITE HIGGS

[Contributor(s): Giuliano Panico, Luigi Delle Rose]

Composite Higgs models (CHMs) provide an elegant solution to the EW hierarchy problem by describing the Higgs boson as a composite state arising from new strongly-coupled dynamics. The most compelling incarnations of this idea identify the Higgs field with a (pseudo-)NGB coming from the spontaneous breaking of a global symmetry that characterizes the new strongly-coupled sector. The interactions with the SM fields (namely the EW and QCD gauge fields and the fermions) are typically obtained through linear mixing (partial compositeness), which guarantees viable gauge and flavor structures.

Early explicit models implementing the composite Higgs paradigm were constructed exploiting the holographic correspondence between a slice of AdS$_5$ and a strongly coupled dynamics in 4 space-time dimensions (see [1528] for a review and for the references to the original works). In these models the Higgs arises as the scalar components of a gauge field propagating in the bulk of AdS$_5$, realizing the so-called gauge-Higgs unification scenario.\footnote{Related models that provide a (partial) solution to the hierarchy problem are the Randall–Sundrum constructions, in which the Higgs is identified with a scalar field localized on the IR brane of an AdS$_5$ slice. In these models, however, the Higgs is not a NGB, thus the gap between the actual Higgs mass and the IR scale (of order few TeV) must be obtained by a small tuning. Some of the cosmological consequences of composite Higgs scenarios that we present in this section, in particular the presence of DM candidates, can be obtained in Randall–Sundrum models as well.} Apart from the Higgs, the various fields propagating in the bulk give rise to towers of heavy...
Kaluza–Klein (KK) modes, analogous to the towers of resonance arising from a strongly-coupled dynamics a la QCD. In generic models, KK towers corresponding to the SM gauge and fermion fields are present.

More modern explicit realizations depart from the extra-dimensional constructions and describe the composite Higgs scenarios through a 4-dimensional EFT. In these models the Higgs dynamics is represented by a non-linear sigma model and heavy composite resonances (analogous to the KK models in the holographic models) can be introduced through a moose structure with a collective breaking mechanism (see [1529] for a review and for the references to the original works).

The NGB nature of the Higgs introduces important changes with respect to the SM predictions. The Higgs potential can be significantly modified, either because of distortions in the Higgs self-couplings or because an extended scalar sector emerges from the strongly-coupled dynamics. Both effects can change the properties of the EWPhT, allowing it to become of the first order (contrary to the SM, where it is a smooth crossover). This possibility, together with new non-linear Higgs couplings that provide additional sources of CPV, could make EWBG a viable option.

Another aspect of CHMs that can have implications for cosmology is the presence of heavy resonances that emerge from the composite sector. The introduction of suitable symmetries can make some of these states stable on cosmological time scales, allowing them to play the role of DM.

In the following we will review these two aspects, providing references to the relevant literature.

8.1. Modified electroweak phase transition and baryogenesis

Deformations of the SM scalar potential are encoded, at leading order, in the Wilson coefficient of the dimension six operator $|H|^6$. This operator can be generated already in the simplest CHM realization, described by the coset $SO(5)/SO(4)$ [1530], in which the Higgs doublet emerges as the only NGB of the strong sector. The role of such operator in providing a first order PhT and efficient out-of-equilibrium dynamics to sustain EWBG was first discussed in [1531, 1532].

As CHMs have been under scrutiny by a strict experimental program, EW precision tests
Higgs coupling measurements [1534], and direct searches for top partners [1535, 1536] have imposed strong bounds on the compositeness scale and thus significantly reduced the impact of the corrections to the Higgs potential. As a consequence, the nature of EWPhT in the minimal scenarios cannot deviate much from the SM one and a first order PhT is not achievable 16.

Successful CHMs exhibiting a first order EWPhT can be realized, instead, with non-minimal symmetry-breaking patterns, the simplest possibility being provided by the coset SO(6)/SO(5) [1538, 1539] in which an extra real scalar $\eta$ emerges as a NGB along with the Higgs doublet. The patterns of EWSB are much richer: the transition can proceed directly from the symmetric phase to the EW broken one or through several steps in which the extra scalar gets a VEV at intermediate temperatures. In the latter case, the barrier between the degenerated minima is generated by the tree-level portal coupling $H^2\eta^2$ and the PhT is found to be much stronger, enhancing the possibility that the associated signal of stochastic gravitational wave spectrum could be observed at future space-based interferometry experiments.

A peculiar feature of CHMs is that the non-linearities introduce additional non-renormalizable Higgs interactions, such as $\eta H\bar{t}_L t_R$, providing new sources of CPV. The operator can be generated by integrating out heavy top partners and its role in the context of EWBG has been recently revisited in [724]. If the symmetry of the strong sector is supplemented by $Z_2$ parity, the CPV in the complex phase of the top mass becomes active only during the EWPhT and turns off at zero temperature. This allows to reproduce the correct amount of matter-antimatter asymmetry without conflicting with the stringent constraints from electron and neutron EDM measurements.

A successful scenario also requires suitable embeddings of the left and right-handed components of the top quark into multiplets of the global symmetry group. These have been explored, for instance, in [1539–1541]. Other cosets have also been considered, such as the SO(7)/SO(6), whose scalar sector is characterized by the Higgs doublet and two additional SM scalar singlets [1542], SU(5)/SO(5) that delivers scalar triplets, SO(7)/G2 that delivers a scalar custodial triplet [1543], and SO(6)/SO(4) × SO(2) providing two Higgs doublets [1544, 1545].

16 Generically the scalar potential of CHM is described by trigonometric functions which lead to multiple degenerate minima that may survive even at finite temperature, such that it must be necessarily lifted to avoid the domain wall problem [1537].
Besides the EWPhT, the strong sector of CHMs undergoes a deconfined/confined phase transition. This was studied in the seminal paper [1546] with a holographic approach in the context of Randall-Sundrum (RS) models at finite temperature stabilized by the Goldberger-Wise mechanism. The models are characterized by two phases: a RS phase at low temperature, describing the SM, and a hot conformal field theory at high temperature. The confinement PhT is found to be first order and the EW one typically happens simultaneously, making the latter first order too. This implies a potentially observable gravitational wave spectrum at LISA [1547] and an efficient production of baryon asymmetry [1548].

In [725, 1549–1551] a PhT at the TeV scale triggered by strongly coupled nearly conformal dynamics has been reanalyzed. These scenarios usually exhibit a long period of supercooling [217]. The properties of the PhT do not depend on the absolute energy scale but only on the amount of supercooling. Nearly conformal dynamics at higher scales [1552, 1553] would lead to a gravitational wave spectrum peaked in the frequency range probed by the LIGO and Virgo interferometers.

In these scenarios, efficient EWBG can also exploit an additional source of CPV obtained from varying Yukawa couplings during the PhT, easily implemented through partial compositeness [1554]. An alternative mechanism for EWBG is also possible and relies on field contents that realize EW symmetry non-restoration at high temperatures [726, 1555–1557]. These models have the advantage that the required extra sources of CPV are decoupled from low-energy processes and thus avoid all the stringent bounds from EDM measurements.

### 8.2. Dark matter candidates

DM candidates can be easily obtained in holographic composite Higgs scenarios by introducing a suitable discrete symmetry that makes some of the massive KK modes stable[17]. Two types of symmetries are often considered in the literature: discrete exchange symmetries that relate different copies of the bulk fields [1558], and geometrical parity symmetries[18] connected to the $S_1/Z_2$ orbifold structure representing the extra spatial dimension (see for instance [1560]).

17 From the 4-dimensional effective description perspective, this corresponds to identify the DM candidate with a heavy massive resonance coming from the composite dynamics.

18 This type of symmetries are analogous to KK parity in universal extra dimension models (see for instance [1559]).
Since all the fields propagating in the bulk of the extra dimension give rise to KK modes, different options are there for obtaining a DM candidate. A natural option is to identify the DM with a $Z_2$-odd vector KK mode associated to a 5-dimensional gauge field. Gauge singlet candidates are easily obtained if the 5-dimensional gauge symmetry of the model (corresponding to the global symmetry of the composite sector) contains U(1) subgroups. This happens, for instance in the minimal models with custodial symmetry, based on the coset $SO(5) \times U(1)_x / SO(4) \times U(1)_x$ [1530]. Models featuring vector DM states have been constructed both on warped space [1561, 1562] and on flat space [1558, 1563].

Stable heavy KK modes can also be obtained for the fermionic fields that propagate in the bulk. In this case natural candidates for DM are present if the lightest odd KK state is neutral. For instance this happens if the neutrino fields propagate in the bulk. Also in this case models on warped space [1560, 1564, 1565] and on flat space [1566, 1567] can be constructed.

A second class of viable DM candidates within the composite Higgs framework arises in non-minimal models featuring an extended scalar sector. The possibility of identifying additional (pseudo-)NGBs with a DM candidate was first pointed out in [1568], where a model based on the symmetry-breaking pattern $SO(6) / SO(5)$ was considered. In addition to the Higgs doublet, this coset gives rise to a singlet, which can be made stable through a suitable $Z_2$ symmetry.

Models of this kind are particularly appealing because they provide a unified origin for the EW scale and the DM mass, explaining the coincidence of the two in the WIMP paradigm. Another interesting aspect is the fact that the Higgs can naturally have small portal couplings with the DM candidate, evading the strong bounds from low-energy detection experiments.

A crucial aspect of these models is the symmetry stabilizing the composite DM. In many models the composite dynamics does not predict such symmetry, which must be introduced by assumption. Interesting counterexamples have been studied in [1569], in which a discrete symmetry (‘DM parity’) arises naturally in a UV completion of the $SU(4) \times SU(4) / SU(4)$ coset, in [1570], where a $U(1)_{\text{dm}}$ symmetry stabilizes a complex singlet, and in [1571], where a $SU(5)$ symmetry is present. However, it must be noticed that in all cases the symmetry that stabilizes the DM must be extended to the interactions between the composite sector and the SM fields by assumption.

A comprehensive study of the cosets giving rise to potential DM candidates has been
Table III. List of cosets that give rise to extended composite Higgs sectors with possible DM candidates. In the third column the NGBs associated to the coset are listed (in SU(2) × U(1)_Y multiplets). The states giving rise (or containing) the DM field are underlined.

<table>
<thead>
<tr>
<th>DM type</th>
<th>Coset symmetry</th>
<th>Scalar sector</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>singlet</td>
<td>SO(6)/SO(5)</td>
<td>2_1/2 + 1_0</td>
<td>1568, 1577, 1578</td>
</tr>
<tr>
<td></td>
<td>SO(7)/SO(6)</td>
<td>2_1/2 + 1_0 + 1_0</td>
<td>1542</td>
</tr>
<tr>
<td></td>
<td>SO(7)/G2</td>
<td>2_1/2 + 1_0 + 1_1</td>
<td>1579, 1580</td>
</tr>
<tr>
<td></td>
<td>SO(7)/SO(5) × SO(2)</td>
<td>2_1/2 + 2_1/2 + 1_0 + 1_0</td>
<td>1581</td>
</tr>
<tr>
<td></td>
<td>SO(5) × U(1)/SO(4)</td>
<td>2_1/2 + 1_0</td>
<td>1542</td>
</tr>
<tr>
<td></td>
<td>SO(6)/SO(4)</td>
<td>2_1/2 + 2_1/2 + 1_0</td>
<td>1572</td>
</tr>
<tr>
<td></td>
<td>SU(7)/SU(6) × U(1)</td>
<td>2_1/2 + 3_{SU(3)} + 1_0 + 1_0</td>
<td>1571</td>
</tr>
<tr>
<td></td>
<td>SU(4) × SU(2) × U(1)/Sp(4) × U(1)</td>
<td>2_1/2 + 1_0 + 1_0 + 1_0 + 1_0</td>
<td>1582</td>
</tr>
<tr>
<td>complex singlet</td>
<td>SO(7)/SO(6)</td>
<td>2_1/2 + 1_0 + 1_0</td>
<td>1570, 1583</td>
</tr>
<tr>
<td>doublet</td>
<td>SO(6)/SO(4) × SO(2)</td>
<td>2_1/2 + 2_1/2</td>
<td>1578</td>
</tr>
<tr>
<td>other</td>
<td>SU(4) × SU(4)/SU(4)</td>
<td>2_1/2 + 2_1/2 + 3_0 + 1_0 + 1_1 + 1_0</td>
<td>1569, 1584</td>
</tr>
<tr>
<td></td>
<td>SU(6)/SO(6)</td>
<td>3_1 + 3_0 + 2_1/2 + 2_1/2 + 2_1/2 + 1_0 + 1_0 + 1_0</td>
<td>1585, 1586</td>
</tr>
<tr>
<td></td>
<td>SU(6)/Sp(6)</td>
<td>2_1/2 + 2_1/2 + 1_0 + 1_0 + 1_0 + 1_0 + 1_1</td>
<td>1587, 1588</td>
</tr>
</tbody>
</table>

Notice that analogous models featuring Goldstone boson DM candidates from an extended Higgs sector have been constructed in the context of Little Higgs theories (see for instance [1573–1575]).
Finally, we mention that in UV completions of the composite Higgs scenarios a la technicolor, DM candidates could also come from technibarions or technimesons that are stable thanks to a U(1) symmetry (see for instance [1584, 1590, 1591]).

9. TWIN HIGGS

[Contributor(s): Saurabh Bansal, Yuhsin Tsai]

The idea of Twin Higgs (TH) [259] provides a unique connection between cosmology and particle physics. The TH models solve the EW hierarchy problem that remains a focal question in particle physics. Furthermore, the new particles predicted by the models lead to new cosmological signatures with length scales that span from the Large Scale Structure and CMB to the galaxy formation and stars.

Conventional solutions to the Higgs little hierarchy problem are expected to produce distinct signals at high energy colliders, due to which they have come under tension with the measurements at the LHC. TH models have been proposed to evade these collider constraints and leave no other observable resonance within reach of the LHC. The original incarnation of the TH theories contain a mirror (“twin”) sector that has exactly the same particle content and interactions as the SM. This replication of the SM, produced through an approximated discrete ($Z_2$) symmetry, ensures the cancellation of problematic quantum corrections to the Higgs mass. The mixing between the Higgs bosons in the two sectors leads to a suppression of the couplings of the Higgs particle to SM states. In order to satisfy these constraints, a mild hierarchy is required between the scale of EWSB in the twin sector, denoted by $\hat{v}$, and the corresponding scale in the SM sector $v$, so that $\hat{v}/v \gtrsim 3$ [264, 1592]. Additionally, the rough requirement of a natural theory, with tuning no worse than 10%, sets an approximate upper bound $\hat{v}/v \lesssim 5$. The bound, however, can be relaxed in non-minimal models that generate a larger $\hat{v}/v$ without additional tuning [1593, 1594]. These bounds give a well-motivated region of $3 \lesssim \hat{v}/v \lesssim 5$; that is to say, the twin particles are about three to five times heavier than their SM counterparts. With this setup, the experimentally-permitted decay of the SM Higgs to twin particles would only lead to missing-energy signature at the colliders. As a result, it would be challenging to attribute such observations uniquely to TH models.

Cosmological observations, on the other hand, provide the strongest probe for this mirror-
TH (MTH) model. Additional radiation from the twin photon and twin neutrinos modify the expansion rate of the universe. Moreover, before the twin protons and twin electrons recombine to form neutral twin atoms, the interaction of twin baryons (now part of DM) with twin photon leads to a suppression in structure formation. The non-free-streaming nature of twin photons before this recombination also generates a phase shift in the CMB power spectrum compared to theories with only free-streaming radiation. As the precision of the cosmological measurements improve, we can keep learning more about a possible twin sector and its role to the Higgs hierarchy problem.

In the simplest MTH model, the presence of twin radiation creates an immediate problem due to the existing bounds on additional radiation from the observed abundance of heavy elements. This problem can be resolved if the temperature of the twin sector is lower than that of the SM. In fact, if the twin and SM sectors decouple (as expected) near a few GeV, then the twin sector will be naturally colder because, at the time of decoupling, more twin species would have already left the thermal bath due to their heavier masses. However, this cooling process is insufficient for the simplest MTH model to satisfy the constraint mentioned above. One could posit additional sources of cooling the twin sector, for example, an asymmetric post-inflationary reheating that preferentially reheats the SM more effectively than the twin sector [271, 272]. Alternative ways to cool the twin sector have also been studied [270, 273, 277, 1595–1597].

Cosmology of the twin sector in the MTH model can be determined by three parameters: (1) $\dot{v}/v$, determines the twin baryon masses, (2) $\Delta N_{\text{eff}}$, determines the twin radiation temperature, and (3) $\dot{r} \equiv \rho_{\text{twin baryon}}/\rho_{\text{DM}}$, fraction of the DM that form twin baryons. Ref. [274, 326] discuss the signals of CMB and matter power spectrum with the presence of MTH sector. Using the lattice results and running of the twin QCD coupling, one can estimate the temperature of twin neutrino decoupling that sets the relative abundance of twin helium and twin hydrogen nuclei. The recombination of the twin hydrogen can be described by the standard calculation but with a heavier electron and different baryon number density, and the depletion of the ionized twin electrons makes the twin photon free-streaming. Compared to the $\Lambda$CDM model where DM has a logarithmic growth of its density perturbation before the time of matter-radiation equality, the twin protons and electrons undergo oscillations before they recombine, leading to a suppressed matter power spectrum for the modes that have already entered the horizon, $k \gtrsim 0.05 \text{ Mpc}^{-1}$. The amount of suppression,
as well as the extra oscillation pattern in the power spectrum, are sensitive to the values of $(\hat{v}/v, \Delta N_{\text{eff}}, \hat{r})$ parameters. For instance, if the temperature of the twin sector is $\approx 40\%$ of the SM temperature ($\Delta N_{\text{eff}} = 0.2$), the current data from Planck 2018 TTTEEE+lowE+lensing+BAO+KV450 allows up to $\approx 5-10\%$ of DM to be the twin baryon with $3 \leq \hat{v}/v \leq 5$ \[^{326}\]. This indicates that the energy density of the twin baryon can be comparable to that of the SM baryon. More importantly, besides providing a solution to the little hierarchy problem, the existence of the twin particles can resolve the $H_0$ and $S_8$ tensions simultaneously if the $S_8$ tension worsens in the future with reduction of systematic uncertainties in the measurements \[^{274, 326}\]. If the twin particles are indeed responsible for these tensions, the MTH model provides a significant improvement over the $\Lambda$CDM model, with the best-fit point at $(\hat{v}/v, \Delta N_{\text{eff}}, \hat{r}) \approx (6, 0.3, 0.2)$ \[^{326}\]. The current data leads to a considerable uncertainty in the determined values of the twin parameters. However, future Euclid and CMB-S4 measurements can substantially improve the precision of these parameters to about $1\%$ for $\hat{r}$ and $10\%$ for $(\hat{v}/v, \Delta N_{\text{eff}})$ \[^{326}\]. Furthermore, the upcoming High-Luminosity LHC can provide an independent measurement of $\hat{v}/v$ to relate the cosmological signatures to the solution of Higgs hierarchy problem.

Besides the CMB and Large Scale Structure, the twin particles can also leave their imprints in galactic structures. During galaxy formation, the radiation pressure from the twin photons slows down the gravitational collapse of the DM halo. However, the emission of twin photons from the charged twin particles dissipate energy from the twin sector and speeds up the gravitational collapse. Therefore, the complicated dynamics of the twin particles may lead to halo formation which may resemble either the standard NFW profile, profile with thermalized DM particles, or a disk-like profile \[^{345}\]. A dedicated N-body simulation of the system is necessary to predict the behavior of the twin halo. The possibility of a disk-like sub-halo is especially interesting since this scenario can be constrained by the GAIA survey \[^{333, 340}\], lensing measurements \[^{341}\], cooling of white dwarfs \[^{348}\], and searches for dark stars \[^{347, 349}\] if the twin photon mixes with the visible photon. Depending on the detailed assumptions of the Yukawa couplings of twin particles, the observed DM density can dominantly come from the neutral twin atoms with an early twin recombination \[^{270, 1595}\]. The atomic twin DM in these scenarios receives constraints from DM self-scattering. The non-trivial halo formation process can generally exist in mirror DM scenarios discussed in Sec. 2.6 whenever the dark atom binding energy is lower than the gravitational energy as-
associated with the halo formation. However, motivated by the Higgs hierarchy problem, the MTH model may provide a more constrained parameter space than the general atomic DM models and allow more precise modeling of the mirror halo today.

The discussion above has mainly focused on the MTH models. However, in order to address the little hierarchy problem, we only need to mirror-symmetrize a subset of the SM particles that dominate the quantum corrections to the Higgs mass [264]. Variations of the TH models that satisfy this weaker symmetry requirement can lead to different realizations of DM. These include, for example, DM that maintains the WIMP miracle [265, 266, 287], ADM that explains the similar baryon and DM energy densities [267, 268], self-interacting DM that address puzzles regarding the small scale structures [269, 275], or DM particles that produce a gamma-ray spectrum similar to the observed galactic center gamma-ray excess [1598].

10. COSMOLOGICAL SOLUTIONS TO NATURALNESS PROBLEMS

[Contributor(s): Tevong You]

The naturalness problem continues to challenge our understanding of EFT. Despite the success of EFT reasoning across many orders of magnitude in scale, it appears to be failing us now. A central principle of EFT is the insensitivity of physics at low energies to the details of what goes on at higher energies. This self-containment of each layer of nature’s descriptions has held across the entire development of physics—and science more generally—thus far. It is therefore puzzling that the Higgs boson mass in the SM and the cosmological constant of General Relativity (GR) violate this fundamental and empirically well-established principle, assuming the SM+GR to be the low-energy descriptions of a more fundamental theory in which these parameters are calculable. Their UV sensitivity means that a small value relative to the EFT cut-off scale can only be obtained by cancelling out large UV contributions coming from the tiniest scales, as if balancing a pencil on its tip by relying on fine tuning at the atomic level to maintain it upright. A consistent possibility, to be sure, but highly implausible.

It has often been argued that the naturalness problem is a mirage. One argument goes that since parameters serve only to relate observables to observables there is no UV sensitivity at this level—parameters are determined by measurement and other observables are then fixed in terms of that measurement, with no quadratic divergencies appearing anywhere.
However, this is no longer the case if a parameter is calculable in the UV, as expected if the Higgs sector originates from a more fundamental description. Then the fine-tuning manifests itself in the structural relations enforced by the theory as heavy physics is decoupled from low energies. Alternatively it could be that the Higgs potential is not calculable, that its mass and couplings are measured quantities with no underlying origin. This is hard to believe, given how tightly knit the remaining structure of the SM is and the evidence so far for nature unifying into ever more rigid frameworks. That this trend should suddenly stop now would be a radical proposal in itself. We would be washing our hands of unfinished business by simply declaring it to be finished. Such drastic action seems rather premature.

Naturalness is therefore an important guiding principle for developing theories beyond the SM, whether it applies in nature or not. This remains the case even as no signs of conventional solutions such as supersymmetry or compositeness or extra dimensions have been discovered around the weak scale. Null results at the LHC may be suggesting a missing component to our reasoning. Indeed, the cosmological constant problem, known as the worst order of magnitude estimate in physics, already indicated that we need something more in our bag of tricks than introducing new symmetries.

Cosmological dynamics in the early universe may provide an alternative solution to the naturalness problem, or at least postpone the appearance of a symmetry-based solution to higher energies without fine tuning. Many proposals have been put forward to address the hierarchy problem of the EW scale [1599–1614] in this way. In one of the earliest attempts, Abbott’s model [1615], the cosmological constant is dynamically determined by a slow-rolling scalar field scanning its value as it evolves down the potential. The evolution stops when Hubble falls below the confinement scale of some strong sector and a periodic potential becomes unsuppressed, acting as barriers that trap the scalar at a small cosmological constant value. Unfortunately, a consequence of the mechanism is a cold and empty universe (though reheating may be possible through a violation of the null energy condition [1616–1618]).

Inspired by Abbott’s model, Graham, Kaplan, and Rajendran (GKR) [1601] used an axion-like field, the so-called relaxion, to scan the Higgs mass. In this case the periodic potential barriers depend on the Higgs VEV which therefore trap the relaxion soon after the transition from unbroken to broken phase. The result is a natural hierarchy between the Higgs mass and the EFT cut-off scale, at the cost of an extremely flat (but technically natural) potential with super-Planckian field excursions that take exponentially long e-foldings.
to scan the necessary field range. The Hubble scale of inflation must also lie below the weak scale for the relaxion to scan the zero-temperature Higgs potential. This leads to a non-generic picture for the UV (for example clockwork or other large field excursion constructions) in which the model is embedded and for the cosmology where the mechanism takes place. Nevertheless the model demonstrates a viable new possibility for resolving EFT naturalness, and many variations have been proposed since, e.g. [1619–1641].

The dependence of the periodic potential on the Higgs VEV implies confinement of a strong sector whose fermions obtain their masses from the Higgs. The original proposal of identifying this sector with QCD is attractive but excluded by predicting too large a strong-CP Θ-angle, in violation of neutron dipole moment bounds. It is difficult to rescue this scenario, for example by introducing thermal effects to modify the barriers before and after scanning, without introducing other problems. Alternatively, a new strong sector of fermions could be responsible. This predicts new physics close to the weak scale, which undermines somewhat the original motivation to explain the lack of such discoveries. On the other hand it provides a potential collider signal distinct from symmetry-based solutions to naturalness that could be used to establish such a mechanism. A more serious objection is the coincidence problem that a new strong sector source of electroweak symmetry breaking just happens to confine close to the weak scale. To avoid this, a periodic potential with a quadratic dependence on the Higgs can decouple the new strong sector to higher scales [1619]. However, this introduces a barrier preventing the relaxion from scanning that requires a second relaxion to relax these barriers. Other realisations of cosmological relaxation of the weak scale use a backreaction of the Higgs VEV elsewhere than the periodic potential.

Particle production by the relaxion as a backreaction mechanism was first considered in Refs. [1624, 1625] (see also Refs. [1629, 1642] for other applications in the context of the original GKR model). There the periodic potential is due to a dark sector and VEV-independent, while the relaxion initially has enough energy to slow-roll over the barriers. When reaching the critical point, a light Higgs VEV triggers particle production to induce extra friction and trap the relaxion. The parameter space of the model of Ref. [1624] was subsequently eroded by constraints from relaxion fragmentation and requiring naturally photophobic relaxion couplings [1643, 1644]. Furthermore the Schwinger effect involving SM fermions had been neglected, which suppresses the necessary dissipation [1644]. The latest particle production model of cosmological relaxation introduced in Ref. [1641] overcomes
these issues by instead making use of the Schwinger effect as a backreaction mechanism.

Relaxion models can be targeted by axion searches and their associated phenomenology in astrophysics, cosmology and gravitational waves \[1640, 1644, 1653\]. They may even be the DM or form a part of it \[1636, 1647, 1648\]. Relaxions have also been linked to mechanisms of leptogenesis \[1632\], baryogenesis \[1636, 1654\], neutrino masses and a Nelson-Barr solution to the strong-CP problem \[1636\]. Cosmological relaxation is not a full solution to the naturalness problem since the EFT cut-off is bounded from above to be anywhere between 100 TeV to $10^{10}$ GeV. However it can be a solution to a little hierarchy problem that allows supersymmetry to be naturally heavy \[1621, 1623\], or for the scale of compositeness in composite Higgs models to be pushed higher \[1627\].

The role of early-universe cosmology in understanding the properties of our universe may be viewed more broadly through the lens of self-organised criticality. The approach of Refs. \[1655–1657\] considers the setting of the string landscape and defines an early-time measure whereby we reside in more accessible vacua as dynamical fluctuations explore the landscape. This corresponds to a self-organisation in the dynamical selection of vacua to reside at a critical point, which could coincide with a small cosmological constant or Higgs mass. Ref. \[1613\] instead focuses on the dynamics of a fluctuating “apeiron” scalar field in a local region of a potential. A new phenomenon, dubbed Self-Organised Localisation (SOL), occurs when the apeiron scanning a parameter triggers a quantum phase transition: the apeiron reaches a stationary volume distribution during inflation that is localised at the critical point between the coexistence of two phases. The parameter value observed in our vacuum today may then violate EFT expectations. SOL may be applied to understand the Higgs potential metastability in the SM, the hierarchy problem of the Higgs mass, or the cosmological constant, amongst other possibilities to be explored.

There may also be signatures of SOL accessible to colliders. In setting the Higgs mass, the prediction for its value is related to the instability scale of the Higgs potential. New physics, for example vector-like fermions, must therefore exist that lowers the instability scale by coupling to the Higgs and contributing negatively to renormalisation group running of its quartic coupling. The existence of only vector-like fermions close to the weak scale may not be entirely ad-hoc; if only the Higgs is dynamically fixed by SOL to be light, then all other scalars and vectors that get their masses from scalars must be heavy, leaving only vector-like fermions as candidates for new physics at the TeV scale. The relation between
stability of the Higgs potential and an upper bound on the Higgs mass has been studied in
Refs. [1613, 1658].

This short review has focused mainly on the motivations for naturalness, relaxion models
and SOL. There are many other cosmological mechanisms for solving naturalness problems
in particle physics, for example: Ref. [1605] inflates patches of the universe with a small
weak scale so that it dominates the volume of the universe by arranging it to coincide
with a maximum in the potential; by contrast, in the sliding naturalness mechanism of
Refs. [1614, 1659], only a small weak scale bounded from above and below can survive
on sufficiently long cosmological timescales after reheating due to a metastable vacuum
disappearing and the patch of the universe crunching for other values of the weak scale.
Reheating can also provide a means of censorship by only allowing reheating for a small
enough Higgs VEV, assuming the existence of a vast number of copies of the SM each
with their own value of the Higgs VEV [1602]. Cosmological dynamics of four-form fluxes
may provide an alternative way to set the EW scale, the cosmological constant, or both
simultaneously [1607, 1660–1662]. Only a few examples have been given here to illustrate
the variety of proposals.

What all these ideas demonstrate is a healthy exploration of alternative paths to BSM
physics. Naturalness remains a central structural problem which must be tackled with
a more open mind than ever. This will be essential for making progress in light of null
results in the search for conventional symmetry-based solutions—an era of so-called post-
naturalness [1663]. Cosmology and the early universe is one such direction that could yet
again demonstrate a rich interplay with particle physics.

ph.CO].


[104] N. D. Birrell and P. C. Davies, Quantum fields in curved space (Cambridge University Press, 1982).


[27] P. A. Zyla et al. (Particle Data Group), PTEP 2020, 083C01 (2020).


137


153


arXiv:2106.13793 [gr-qc].


Acknowledgments


[1095] A. A. Garcia, K. Bondarenko, S. Ploeckinger, J. Pradler, and A. Sokolenko, JCAP 10, 011


A. Fattahi, J. F. Navarro, T. Sawala, C. S. Frenk, L. V. Sales, K. Oman, M. Schaller, and


[1223] O. Sameie, M. Boylan-Kolchin, R. Sanderson, D. Vargya, P. F. Hopkins, A. Wetzel,


196

197


