



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

EXZELLENZCLUSTER
QUANTUM UNIVERSE



EXPLORING THE DARK UNIVERSE WITH GRAVITATIONAL WAVE BACKGROUNDS

Dissertation

zur Erlangung des Doktorgrades an der Fakultät für Mathematik,
Informatik und Naturwissenschaften

Fachbereich Physik der Universität Hamburg

vorgelegt von

CARLO TASILLO

Hamburg
2024

Exploring the dark universe with gravitational wave backgrounds
© Carlo Tasillo 2024

GUTACHTER:INNEN DER DISSERTATION:

Dr. Kai Schmidt-Hoberg
Prof. Dr. Géraldine Servant

ZUSAMMENSETZUNG DER PRÜFUNGSKOMMISSION:

Dr. Kai Schmidt-Hoberg
Prof. Dr. Géraldine Servant
Dr. Thomas Konstandin
Prof. Dr. Jochen Liske
Prof. Dr. Oliver Gerberding

VORSITZENDER DER PRÜFUNGSKOMMISSION:

Prof. Dr. Jochen Liske

DATUM DER DISPUTATION:

14. Oktober 2024

VORSITZENDER DES FACH-PROMOTIONSausschusses PHYSIK:

Prof. Dr. Markus Drescher

LEITER DES FACHBEREICH PHYSIK:

Prof. Dr. Wolfgang J. Parak

DEKAN DER FAKULTÄT MIN:

Prof. Dr.-Ing. Norbert Ritter

ABSTRACT

Gravitational waves have recently emerged as a novel messenger in astronomy, providing profound insights into the evolution of the cosmos: Last year, several pulsar timing arrays (PTAs) announced the detection of a gravitational wave background (GWB) at nHz frequencies, marking a significant milestone in this field. This discovery can be interpreted as a signal emitted by the inspiral of supermassive black holes during hierarchical structure formation. More excitingly, the signal's origin could alternatively predate recombination and Big Bang nucleosynthesis. Such an explanation would require new physics in order to explain the strong dynamics in the primordial plasma above the MeV temperature scale necessary to emit those gravitational waves.

In this thesis, we explore the possibility that PTAs could have detected a cosmological phase transition within a dark sector. A better fit to the data than that for supermassive black hole binaries can be found if the dark sector hosting the transition decays before the onset of nucleosynthesis.

Furthermore, it is plausible that PTAs have instead observed a GWB emitted by the inspiral of supermassive *primordial* black holes. Our research indicates that these primordial black holes can account for the observed signal only if they are initially clustered rather than homogeneously distributed.

Looking towards future gravitational wave observatories, we examine the formidable opportunity offered by LISA, that is sensitive to mHz frequencies, to probe GWBs emitted at temperatures of the primordial plasma around the 100 GeV scale, aligning with the freeze-out epoch of a WIMP dark matter candidate. Our findings suggest that if LISA detects such a cosmological GWB, it would hint towards a dark sector phase transition producing the relic dark matter abundance, consistent with cosmological and astrophysical observations.

Our results underscore the significance of gravitational wave observations in unveiling the dynamics of the early universe and provide new pathways for understanding the fundamental nature of dark matter and the cosmological history of the universe.

ZUSAMMENFASSUNG

Gravitationswellen haben sich jüngst als neuartige Boten in der Astronomie herausgestellt und liefern tiefgreifende Einblicke in die Entwicklung des Kosmos: Letztes Jahr kündigten mehrere PTAs die Entdeckung eines Gravitationswellenhintergrunds bei nHz-Frequenzen an, was einen bedeutenden Meilenstein in diesem Bereich darstellt. Diese Entdeckung kann als Signal interpretiert werden, das durch die Verschmelzung von supermassiven Schwarzen Löchern während der hierarchischen Strukturentstehung emittiert wurde. Noch spannender ist die Möglichkeit, dass der Ursprung des Signals alternativ vor der Rekombination und der Urknall-Nukleosynthese liegt. Eine solche Erklärung würde neue Physik erfordern, um die starke Dynamik im primordialen Plasma über der MeV-Temperaturskala zu erklären, die zur Emission von diesen Gravitationswellen notwendig ist.

In dieser Dissertation untersuchen wir die Möglichkeit, dass PTAs einen kosmologischen Phasenübergang innerhalb eines dunklen Sektors detektiert haben könnten. Eine bessere Übereinstimmung mit den Daten als bei supermassiven Schwarzen-Loch-Binärsystemen kann gefunden werden, wenn der dunkle Sektor, der den Übergang beherbergt, vor dem Beginn der Nukleosynthese zerfällt.

Darüber hinaus ist es plausibel, dass PTAs stattdessen einen Gravitationswellenhintergrund beobachtet haben, der durch die Verschmelzung von supermassiven *primordialen* Schwarzen Löchern emittiert wurde. Unsere Forschung zeigt, dass diese primordialen Schwarzen Löcher das beobachtete Signal nur dann erklären können, wenn sie anfangs geklumpt und nicht homogen verteilt sind.

Mit Blick auf zukünftige Gravitationswellen-Observatorien untersuchen wir die interessante Koinzidenz, dass LISA für mHz-Frequenzen empfindlich sein wird, die Gravitationswellen-Hintergründen entsprechen, die bei Temperaturen des primordialen Plasmas um die 100 GeV-Skala emittiert werden und mit der Freeze-out-Epoche eines WIMP-Dunkle-Materie-Kandidaten übereinstimmen. Unsere Ergebnisse legen nahe, dass, wenn LISA einen solchen kosmologischen Gravitationswellenhintergrund detektiert, dies auf einen Phasenübergang in einem dunklen Sektor hinweisen würde, der die Dunkle Materie-Menge produziert, die mit kosmologischen und astrophysikalischen Beobachtungen übereinstimmt.

Unsere Ergebnisse unterstreichen die Bedeutung von Gravitationswellenbeobachtungen für die Enthüllung der Dynamik des frühen Universums und bieten neue Wege zum Verständnis der grundlegenden Natur der Dunklen Materie und der kosmologischen Geschichte des Universums.

*Habt ihr euch mal überlegt,
wenn die Schwerkraft Runden dreht,
als Welle durch die Erde schwebt,
gemächlich durch die Häuser zieht,
Bratkartoffeln brotscheln riecht:
Geht sie dann bei Freunden ein,
setzt sich an das Feuerlein
und Schwerkraft lässt mal Schwerkraft sein –
wie schmecht ihr dann der rote Wein?*

— Valerie Weber

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to *Kai Schmidt-Hoberg* for offering me the opportunity to pursue my PhD and for his outstanding supervision throughout these years. Your approach to supervision—providing ample freedom concerning research topics and working hours—allowed me to grow as an independent researcher. You made the start of my PhD journey easy, even during the challenging times of the COVID-19 pandemic, and set a remarkable example of how to balance family life and a demanding career. Your guidance during my postdoc decisions was invaluable, and for all of this, I am truly grateful.

I am also thankful to *Géraldine Servant* for her co-supervision and to *Thomas Konstandin* for his mentoring, both professionally and personally. The insightful discussions and a memorable trip to Reeperbahn were a highlight of my time in Hamburg. I would also like to thank *Jochen Liske* and *Oliver Gerberding* for agreeing to join my disputation committee.

A special mention goes to my collaborators who have greatly enriched my research experience. *Felix Kahlhöfer*, your supervision and collaboration have extended far beyond the limits of my Master's thesis, and I am deeply thankful for your continued support, including writing reference letters and inviting me to Karlsruhe. *Torsten Bringmann*, your insights into statistical data analysis and our numerous collaborations have been a cornerstone of my research. Your invitation to Oslo and your invaluable support for my postdoc application were important steps in my academic journey. *Pedro Schwaller*, thanks a lot not only for sharing your insights on the latest movements within the PTA community, but also for amusing bouldering sessions in Mainz and Munich.

I am particularly grateful to *Frederik Depta* for being a valuable collaborator throughout my PhD. Your life advice during the challenging period of postdoc applications was deeply appreciated. I would also like to thank *Jonas Matuszak* and *Tomás Gonzalo* for our close collaboration and the many fruitful discussions we have had. Additionally, I appreciated the engaging conversations with *Joachim Kopp* and *Yann Gouttenoire* at Fermilab and CERN which have greatly influenced my work.

I am grateful to the Quantum Universe cluster of excellence for giving me the opportunity to travel the world during my PhD, and to the Fermilab theory group for their hospitality during my visit. The latter was made possible thanks to an ASYMMETRY grant and the Horizon Europe Programme Staff Exchange project. I also want to thank the Swedish Collegium for Advanced Study for allowing me to attend a conference in Uppsala, even when my own funding was limited, ultimately helping me find my future postdoc position.

I am thankful to the Studienstiftung des deutschen Volkes and Cusanuswerk for supporting me during my Bachelor's and Master's studies. Your impact on my journey as researcher and part of society has been profound and positive. My sincere thanks also go out to my former teachers, *Christoph Schuck*, *Florian Peter*, *Thomas Gerritzma*, and *Heiko Nüllmann*, who fueled my insatiable curiosity in school and helped me find my place in physics with their guidance.

I would also like to express my gratitude for the unforgettable experiences I had while participating in the Deutsche SchülerAkademie in 2023 and 2024. Special thanks to *Simon Schwarz* for co-teaching our “Big Bang Math” course. You showed me how to convey complex concepts in a way that students could truly understand. I am also immensely grateful to our students, whose countless insightful questions constantly challenged the limits of my own comprehension and, in turn, helped me become a better physicist. The time spent at these academies was as enriching for me as I hope it was for those I had the privilege of teaching.

A big thank you to my academic brother *Jonas*, my office mates *Felix* and *Andrii*, as well as to the former group members *Philine*, *Apratim*, *Aleix*, *Jeremy*, *Lorenzo*, and *Philipp* for making the office a welcoming place. A long night in the office after visiting “Samarkand”, when we “absolutely needed to finish writing a paper”, will always be a special memory.

I am thankful to *Henda Mansour*, as well as *Jonas Frerick* and *Jonas Matuszak* for proofreading parts of this thesis. I really appreciate the time and effort you put into correcting my repeated errors and the many occasions when you saved me from writing nonsense.

I am especially grateful to my tomato-European friends in the “danger zone”—*Margherita*, *Sara*, *Davide*, *Lorenzo*, and *Alain*. Vi voglio tanto bene, raga.

Finally, I would like to thank my family. *Sandro*, ich danke dir von Herzen für deine brüderliche Hilfe aus der Ferne – nicht nur als ich kopflos eine ganze Nacht in Rom ums Kolosseum gelaufen bin. *Mama* und *Papa*, eure bedingungslose Liebe und Unterstützung haben mich auf meinem gesamten bisherigen Lebensweg begleitet. Diese Thesis ist genauso auch euer Verdienst.

PUBLICATIONS

This thesis is based on the following publications:

- [1] T. Bringmann, P. F. Depta, T. Konstandin, K. Schmidt-Hoberg, and C. Tasillo, *Does NANOGrav observe a dark sector phase transition?*, *JCAP* **11** (2023) 053, [[2306.09411](#)]
- [2] T. Bringmann, T. E. Gonzalo, F. Kahlhoefer, J. Matuszak, and C. Tasillo, *Hunting WIMPs with LISA: correlating dark matter and gravitational wave signals*, *JCAP* **05** (2024) 065, [[2311.06346](#)]
- [3] P. F. Depta, K. Schmidt-Hoberg, P. Schwaller, and C. Tasillo, *Do pulsar timing arrays observe merging primordial black holes?*, [[2306.17836](#)]

The following publication was completed before the start of the doctorate and will therefore not be discussed in detail. Instead, its results are incorporated into the introductory chapters of this thesis, as both projects [1] and [2] are based on it:

- [4] F. Ertas, F. Kahlhoefer, and C. Tasillo, *Turn up the volume: listening to phase transitions in hot dark sectors*, *JCAP* **02** (2022), no. 02 014, [[2109.06208](#)]

CONTENTS

1	Introduction	1
2	Gravitational wave cosmology	5
2.1	Chronology of our universe	5
2.2	Open questions	10
2.3	The homogeneous universe	11
2.3.1	The FLRW metric	12
2.3.2	Thermodynamics of the primordial plasma	14
2.3.3	Dark and visible sectors	17
2.3.4	Going out of equilibrium	21
2.3.5	The freeze-out of dark matter	23
2.3.6	The Big Bang nucleosynthesis	26
2.4	Cosmological gravitational wave backgrounds	29
2.4.1	Gravitational waves in vacuum	31
2.4.2	Helicity decomposition of metric perturbations	34
2.4.3	Gravitational waves in a curved background	37
2.4.4	Gravitational waves in an expanding background	39
2.4.5	Equation of motion for a gravitational wave	41
2.4.6	Stochastic GW signals from the early cosmos	43
2.4.7	The spectrum of primordial gravitational waves	45
2.4.8	The redshift of gravitational wave spectra	48
2.4.9	Primordial GWB searches	49
3	Pulsar timing arrays	53
3.1	A short history of pulsar timing	53
3.2	Gravitational wave effects on a single pulsar's timing	55
3.3	Effect of a GWB on an array of pulsars	57
3.4	The PTA likelihood	60
3.5	Evidence for a gravitational wave background	65
3.6	Supermassive black hole binaries	68
3.6.1	Binary evolution and the final parsec problem	68
3.6.2	GWBs from inspiraling binaries	69
4	Dark sector phase transitions	73
4.1	Finite-temperature effects in QFT	73
4.1.1	A quantum harmonic oscillator in a thermal bath	73
4.1.2	The Kubo-Martin-Schwinger relation	74
4.1.3	The effective potential of an abelian gauge theory	75
4.2	Bubble nucleation and percolation	83
4.3	Gravitational waves from dark sector phase transitions	88
4.3.1	Bubble collisions and sound waves	90
4.3.2	The dilution of GWs due to a dark sector decay	92
5	Do PTAs observe a dark sector phase transition?	93
5.1	Introduction	93
5.2	Computation of GW spectra	95
5.3	Pulsar timing array data analysis	97

5.4	Cosmological constraints	100
5.4.1	Stable dark sectors	100
5.4.2	Decaying dark sectors	101
5.5	Results	102
5.5.1	Stable dark sector phase transitions	102
5.5.2	Decaying dark sector phase transitions	105
5.5.3	Comparison with SMBHBs and later data sets .	107
5.6	Conclusions	110
5.A	Details on the cosmological likelihood	113
5.A.1	Stable dark sector	113
5.A.2	Decaying dark sector	114
5.B	Posterior distribution of GWB spectra	118
5.C	Details on the calculation of Bayes factors	119
5.C.1	The product space method	119
5.C.2	Uncertainties of the computed Bayes factors . .	121
5.C.3	Relating Bayes factors to <i>p</i> -values and <i>Z</i> -scores .	122
5.C.4	Influence of the prior choice on the Bayes factor .	122
5.C.5	Influence of priors on the credibility of a DSPT .	124
5.C.6	Priors for the Bayesian model comparison	126
6	Hunting WIMPs with LISA	127
6.1	Introduction	127
6.2	Dark sector model	130
6.2.1	The effective potential	130
6.2.2	Properties of the phase transition	133
6.2.3	The gravitational wave spectrum	133
6.3	The dark sector relic density	136
6.4	Thermalization of the two sectors	140
6.4.1	The dark sector temperature	140
6.4.2	Thermalization within the dark sector	142
6.4.3	Thermalization of the dark and visible sector .	145
6.5	Gravitational waves from hot dark sectors	148
6.6	Conclusions	151
6.A	Dark bubble walls	155
6.B	Detail on thermalization and freeze-out	157
6.B.1	The thermal mixing angle	157
6.B.2	The Boltzmann equation for entropy transfer .	158
6.B.3	Annihilation cross sections	161
7	Do PTAs observe inspiraling primordial black holes?	163
7.1	Introduction	163
7.2	Gravitational wave signal	164
7.3	Expected number of binaries	168
7.4	PBH production and constraints	170
7.5	PTA data analysis	170
7.6	Results	171
7.7	Discussion and conclusions	173
8	Summary	175

ACRONYMS

BBN	Big Bang nucleosynthesis
BBO	Big Bang Observatory
CMB	cosmic microwave background
CPTA	Chinese Pulsar Timing Array
CURN	common uncorrelated red noise
DM	dark matter
DMGP	dispersion measure as a Gaussian process
DMX	dispersion measure extension of the timing ephemeris
dof	degree of freedom
DS	dark sector
DSPT	dark sector phase transition
ET	Einstein Telescope
EPTA	European Pulsar Timing Array
EWPT	electroweak phase transition
FLRW	Friedmann–Lemaître–Robertson–Walker
FOPT	first-order phase transition
GR	General Relativity
GW	gravitational wave
GWB	gravitational wave background
HD	Hellings-Downs
InPTA	Indian Pulsar Timing Array
IPTA	International Pulsar Timing Array
IR	infrared
JWST	James Webb Space Telescope
Λ CDM	Λ cold dark matter
LIGO	Laser Interferometer Gravitational-Wave Observatory
LISA	Laser Interferometer Space Antenna
LO	leading-order
LVK	LIGO-VIRGO-KAGRA
MCMC	Markov chain-Monte Carlo
NANOGrav	North American Nanohertz Observatory for Gravitational Waves
no-CURN	no common uncorrelated red noise
NLO	next-to-leading-order
PBH	primordial black hole
PLI	power-law integrated
PPTA	Parkes Pulsar Timing Array
PT	phase transition
PTA	pulsar timing array
QCD	quantum chromodynamics
QFT	quantum field theory
SMBHB	supermassive black hole binary

SM	Standard Model
SNR	signal-to-noise ratio
SVT	scalar, vector and tensor
TOA	time of arrival
TT	transverse-traceless
UV	ultraviolet
vev	vacuum expectation value
WIMP	weakly interacting massive particle

NOTATION

In this thesis, Einstein's summation convention $\sum_\mu a_\mu b^\mu \equiv a_\mu b^\mu$ and the Dirac slash notation $\not{k} \equiv k^\mu \gamma_\mu$ are used. If not stated otherwise, natural units in which $c = \hbar = k_B = 1$ are employed. Further, space-time indices are written as Greek letters, while spatial vectors are printed in boldface (\mathbf{x}) with components denoted by Latin indices (x_i). The flat Minkowski metric is chosen as $\eta_{\mu\nu} \equiv \text{diag}(-1, +1, +1, +1)$, where the 0-component corresponds to the time variable (i.e. $x^\mu = (x^0, \mathbf{x})$ with $x^0 = t$). Partial derivatives with respect to a generic variable x are abbreviated as ∂_x ; space-time derivatives read $\partial_\mu = \frac{\partial}{\partial x^\mu} = (\partial_t, \partial_i)$, while total derivatives with respect to time are denoted by $\dot{f}(t) = \frac{df}{dt}$. A particle's four-momentum is defined as $p^\mu = (E, \mathbf{p})$, such that $p_\mu x^\mu = -Et + \mathbf{p} \cdot \mathbf{x}$ and $d^4 p = dE d^3 p$.

The Einstein tensor is given by $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ with the Ricci scalar $R \equiv R^\mu{}_\mu$ and the Ricci tensor $R_{\mu\nu} \equiv R^\gamma{}_{\mu\gamma\nu}$, which is, in turn, defined over the Riemann tensor

$$R^\mu{}_{\nu\rho\sigma} \equiv \partial_\rho \Gamma^\mu_{\nu\sigma} - \partial_\sigma \Gamma^\mu_{\nu\rho} + \Gamma^\mu_{\alpha\rho} \Gamma^\alpha_{\nu\sigma} - \Gamma^\mu_{\alpha\sigma} \Gamma^\alpha_{\nu\rho}.$$

The Christoffel symbols

$$\Gamma^\rho_{\mu\nu} \equiv \frac{1}{2}g^{\rho\sigma}(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu})$$

can be calculated as space-time derivatives of the metric $g_{\mu\nu}$. To simplify the Einstein equations, the reduced Planck mass $m_{\text{Pl}} = (8\pi G)^{-1/2} = 2.4 \cdot 10^{18} \text{ GeV}$ is employed, where G is Newton's gravitational constant. By default, energies, temperatures, masses, and momenta are all given in units of $\text{eV} = 1.6 \cdot 10^{-19} \text{ J}$, whereas distances and time intervals are expressed in inverse energy units.

1 | INTRODUCTION

*Ground Control to Major Tom:
Take your protein pills and put your helmet on
— Space Oddity by DAVID BOWIE*

Gravitational waves were first detected by the LIGO collaboration in September 2015 [5], a century after they were first predicted by Albert Einstein in 1915 [6]. This groundbreaking discovery marked a significant breakthrough in physics and astronomy in particular. In 2023, the field saw another leap forward when pulsar timing arrays (PTAs) announced the first detection of a gravitational wave background (GWB) [7–10]. Until now, gravitational waves (GWs) have primarily been used as probes of astrophysical phenomena, such as black hole and neutron star mergers. However, with the PTA detection, the field of GW cosmology is emerging, opening new avenues for understanding the expansion history and matter content of our universe [11].

The impact of the PTA discovery on cosmology can be compared to the first observation of the cosmic microwave background (CMB). Only the following measurements of the CMB anisotropies made precision cosmology possible and for the first time allowed testable statements about the universe up to temperatures of a billion Kelvin (i.e., the MeV scale) [12]. This period corresponds to the epoch of Big Bang nucleosynthesis (BBN), which occurred shortly after the decoupling of neutrinos. The latter happened at a point in time, when the interactions between neutrinos and other particles of the Standard Model (SM) of particle physics could no longer be sustained due to the decreasing temperature of the primordial plasma. More quantitatively, this point in time can be determined by comparing the weak interaction rate $\Gamma_{\text{weak}}(T)$ with the Hubble expansion rate $H(T)$ [13]:

$$\frac{\Gamma_{\text{weak}}}{H} \sim \frac{G_F^2 T^5}{T^2/m_{\text{Pl}}} \sim \left(\frac{T}{\text{MeV}}\right)^3, \quad (1.1)$$

where $G_F = 1.17 \cdot 10^{-5} \text{ GeV}^{-2}$ is the Fermi constant. We find that, as soon as the temperature of the primordial plasma dropped below the MeV scale ($T \lesssim \mathcal{O}(\text{MeV})$), neutrinos could no longer interact with the surrounding SM particles and henceforth evolved as a decoupled, free-streaming species.

In fact, the same argument can be repeated for GWs [11, 14]: When the gravitational scattering rate

$$\frac{\Gamma_{\text{gw}}}{H} \sim \frac{G_N^2 T^5}{T^2/m_{\text{Pl}}} \sim \left(\frac{T}{m_{\text{Pl}}}\right)^3 \quad (1.2)$$

The dawn of GW cosmology

Comparing neutrino decoupling ...

*... to the
decoupling of GWs*

dropped below the Hubble rate, the plasma was sufficiently dilute and cold enough in order to let GWs propagate freely. From this simple scaling argument we can infer that GWs decoupled already at extremely high temperatures around the Planck scale, namely around a temperature of 10^{32} K. Consequently, unlike other messengers from the early cosmos like photons, which only decoupled at a temperature of roughly 3000 K during recombination [13], GWs could free-stream since the Planck epoch: GWs only redshifted to lower frequencies and weaker amplitudes, preserving their information about the very early cosmos up to the present day. Now, it is on us to find and interpret these signals.

However, the sheer weakness of the gravitational interaction set by Newton's constant $G_N = 6.71 \cdot 10^{-39} \text{ GeV}^{-2}$ is a mixed blessing for cosmologists: On the one hand, in principle it allows the observation of phenomena up to the Planck epoch as shown above. On the other hand, only the most extreme events in which incredible amounts of mass and energy move close to the speed of light, can emit GWs which are strong enough for us to observe them. Following the recent PTA measurements, now we are confronted with the burning question what tremendous events could have sourced the novel PTA signal in the nHz frequency band.

The leading hypothesis for the signal's origin is an astrophysical GWB: During the hierarchical structure formation, galaxies merged forming larger galaxies and clusters [15]. During that process, supermassive black holes residing at the center of galaxies formed bound systems which, assuming the existence of sufficiently effective mechanisms for dissipating binding energy, later inspiraled and merged. Their inspiral emits strong GW signals in the nHz frequency band. Still, there are ongoing debates whether such a mechanism for dissipating a sufficient amount of binding energy is realized in nature [16–24] and whether the amplitude and spectral tilt of the observed GW spectrum can be explained by realistic populations of *astrophysical* supermassive black holes [15]. In this thesis, we explore two alternative, cosmological mechanisms in detail: Dark sector phase transitions and the inspiral of *primordial* black holes.

*Supermassive black
hole binaries*

The study of dark sector phase transitions is not only intriguing as they offer an explanation to the PTA signal. Instead, it provides new insights into the yet obscure realms of our universe, whose energy content is dominated by dark energy and cold dark matter. In the following chapter 2 we will see that the history of our universe can be understood as a series of consecutive phase transitions. As 95% of our universe are referred to as *dark* and are hence assumed to interact possibly only gravitationally with other matter, it is reasonable to further assume that also a dark sector (to be defined properly in section 2.3.3) featured a phase transition in the early universe. Such a phase transition could have not only given rise to dark matter but also to a still-observable GWB [25–27].

*Dark sector phase
transitions*

*G_N: A mixed
blessing*

Primordial black holes on the other hand, while being an interesting dark matter candidate in their own right [28], could have played a decisive role in the early cosmos: There are observations of quasars at high redshifts ($z > 6$, corresponding to a time less than a billion years after the end of inflation) which challenge conventional models of black hole growth [29–33]. A population of primordial black holes (PBHs) formed in the early universe, could have provided a seed mechanism for these early supermassive black holes in the center of the first galaxies. Moreover, these PBHs could have played an important role in the formation of large-scale structure by acting as starting points of galaxy formation [34]. Finally, the formation of PBHs in the early universe through the collapse of large density fluctuations is directly linked to strong dynamics, e.g. coming from cosmological phase transitions or domain wall networks [35, 36]. The study of GWBs, phase transitions as well as PBH formation together with their eventual inspiral and merger is hence full of intriguing interconnections. This motivates us to also consider inspiraling PBHs as an origin for the recently observed PTA signal in this thesis.

After setting the stage for this thesis in an introduction to GW cosmology in chapter 2, followed by a review of the latest PTA results in chapter 3, and an in-depth description on how to calculate GWB predictions for a given dark sector phase transition in chapter 4, we discuss three key questions of this thesis: First, in chapter 5, we aim at answering the question what the odds for a dark sector phase transition explanation of the novel nHz signal are. Second, going beyond the nHz frequency range in chapter 6, we ask what a similar detection of a GWB at the future GW observatory LISA could teach us about the production of WIMP dark matter in the early universe through a freeze-out triggered by a dark sector phase transition. The third question we want to answer in this thesis, in chapter 7, concerns the conditions under which the PTA signal can be interpreted as a GWB from inspiraling primordial black holes. We summarize our results and conclude this thesis in chapter 8.

Primordial black holes

Outline of this thesis

2

GRAVITATIONAL WAVE COSMOLOGY

I became an astronomer because I could not imagine living on Earth and not trying to understand how the Universe works.

— Vera Rubin

In this chapter we want to introduce the field of GW cosmology. We start by giving a brief review of the main events that shaped the universe we live in, in section 2.1. In section 2.2 we then aim at giving a concise overview of the successes the concordance model of cosmology achieved and which unsolved puzzles it leaves us with. The following section 2.3 provides a summary of our description of the expansion of space and its effects on the primordial plasma. Section 2.4 then introduces the concept of GWs in both flat space and a curved background. We will find that GWs stemming from sources active in the early cosmos are necessarily of stochastic nature today and can leave an observable imprint on the observables of precision cosmology. This chapter will end with a discussion of the existing limits on cosmological GWBs, stemming from precision cosmology and direct (non-)observations of gravitational radiation. Large parts of this chapter are based on the two famous books by Michele Maggiore [14, 37] and the detailed review [11] by Chiara Caprini and Dani Figueroa. The initial sections on the chronology and evolution of our universe on large scales build up on the excellent books by Rocky Kolb and Michael Turner [38], as well as Daniel Baumann [13].

Outline of this chapter

2.1 CHRONOLOGY OF OUR UNIVERSE

The best model at our disposal to describe the evolution of the universe is the concordance model. The model can be viewed as a consequence of the two most fundamental theories human minds came up with so far: Albert Einstein’s theory of General Relativity (GR), describing gravity, and quantum field theory (QFT), describing all other interactions. The QFT that made the mathematical formulation of the concordance model possible is the Standard Model (SM) of particle physics. In order to make observations fit with theoretical predictions, several extensions of the SM needed to be introduced, most prominently dark energy (Λ) and cold dark matter (CDM). These necessary additions to the SM give rise to the alternative, more technical name Λ cold dark matter (Λ CDM) model for the concordance model—and to two yet unsolved mysteries of our universe. The objective of this section is to give a chronological

Λ CDM

overview of the events which shaped our universe and are described by Λ CDM. A depiction of this timeline can be found in fig. 2.1.

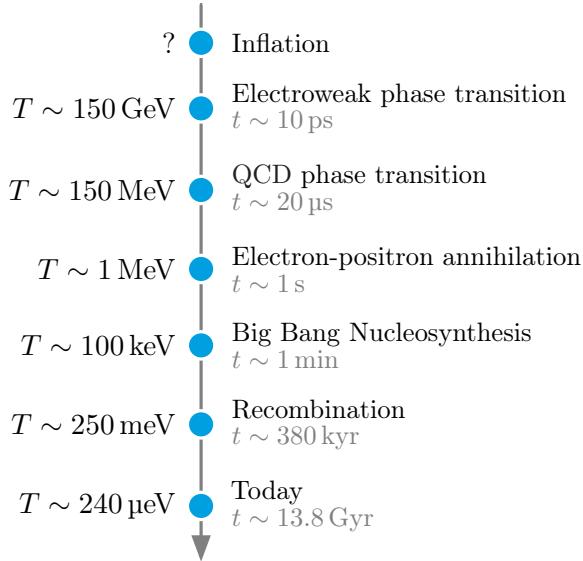


Figure 2.1: A timeline of important events in Λ CDM cosmology.

The Big Bang

The central observation that made today's theory of cosmology possible is that space expands. Our universe must have hence started in a hot and dense state, which is colloquially referred to as the Big Bang. Due to the mathematical limitations of combining GR and QFT in a single, agreed-on, more fundamental theory of everything for describing the interactions of matter, only time scales larger than a Planck time ($t_{\text{Pl}} \simeq 5 \cdot 10^{-44} \text{ s}$) and length scales larger than a Planck length ($\ell_{\text{Pl}} \simeq 2 \cdot 10^{-35} \text{ m}$) can be described reliably. In particular, this forbids any clarifying statements about the initial singularity, which appears if one extrapolates from today's 13.8 billion years old universe back to the state where its energy density exceeded the Planck scale.

Inflation

The earliest event Λ CDM postulates is cosmic inflation. During this process, which lasted at least 10^{-33} s , space expanded by a factor of at least 10^{26} . Inflation is thought to be triggered by a phase transition (PT) of the so-called inflaton field, which also extends the SM. Due to the inflaton's high vacuum energy density, space expanded exponentially fast up to the point when the yet cold universe reheated to some temperature below 10^{15} GeV . During this period of reheating, decays of the inflaton field rapidly filled up the empty space with a thermal bath of SM particles. Due to the strong inflationary period, the plasma is thought to be highly homogeneous, with small inhomogeneities coming from quantum fluctuations of the inflaton field which became macroscopic throughout inflation. These imperfections eventually led to the accumulation of dark matter, acting as a seed for baryonic ("visible") matter over-densities, which much later evolved into stars, galaxies and larger structures. Additionally, also relic GWs were produced both through inflation and the subsequent reheating. So far, only the existence of density perturbations with an almost scale-invariant power spectrum has been

observed, whereas the GWs emitted at these early times have not yet been detected [12].

The next event, which we can be certain occurred, is baryogenesis. During this process, whose precise underlying mechanism is yet unclear, the asymmetry between matter and antimatter has been produced. An often hypothesized connection is with the electroweak phase transition (EWPT), happening at temperatures of around 100 GeV. During this PT, the Higgs field of the SM obtained its vacuum expectation value, thus giving mass to itself, gauge bosons and fermions. In the SM, this PT is a cross-over, i.e. it happens at the same point in time throughout the primordial plasma. There exist several extensions of the SM which render the EWPT first-order, such that it succeeds through the nucleation of bubbles of the new vacuum state, which then expand, collide and perturb the primordial plasma. Under certain circumstances, the present-day asymmetry between the abundance of matter and antimatter could have been produced in the vicinity of these bubbles [13]. In case the EWPT is first-order, GWs with mHz frequencies are expected to have been emitted, lying well within the range where the Laser Interferometer Space Antenna (LISA) will be most sensitive. Hopefully, LISA will hence be able to shed light on the mechanism through which the electroweak symmetry was broken in the early universe.

The electroweak phase transition

Most particle species of the primordial plasma became massive as a result of the EWPT, the only exceptions being photons, gluons and possibly neutrinos remaining massless.¹ When those massive particles can no longer be produced through thermal processes, because their rest-mass can no longer be reached by the characteristic energy scale of particle interactions set by the decreasing temperature, they typically freeze out of the thermal plasma or become Boltzmann-suppressed. In particular, the thermal freeze-out mechanism is often thought to be the origin of dark matter (DM) in our universe, as it naturally gives rise to a relic abundance of particles which at most interact weakly with SM particles, matching astrophysical observations. The precise point in time of DM freeze-out is a model-dependent question. One of the most studied DM candidates, the so-called weakly interacting massive particle (WIMP), is expected to freeze out around the GeV–TeV scale.

Freeze-out of DM

The next noteworthy event happening in the early universe plasma is two more PTs happening at once: Deconfinement and the chiral PT are two intertwined PTs in which quarks and gluons became confined within hadrons, and quarks formed non-zero chiral condensates, respectively. Both events are usually subsumed and referred to as the quantum chromodynamics (QCD) phase transition, as in the SM they occur simultaneously at a temperature of around 150 MeV. Both transitions are cross-overs in the SM, meaning that the process happens at once throughout the universe and no GWs are expected to be emitted during

The QCD phase transition(s)

¹ The SM does not allow for the generation of neutrino masses. From the observation of neutrino oscillations we can, however, infer that at least two neutrino families must be massive [13, 39].

the PT. There exist extensions of the SM which predict one of the two transitions to succeed through the nucleation and expansion of bubbles. The GWs emitted through this process could have already been observed with PTAs or be probed in the μHz -band [40].

Around the MeV scale several events then happened in quick succession: First, neutrinos decoupled from the thermal bath, which was henceforth composed of neutrons, protons, electrons, positrons, photons and a still negligible amount of light nuclei. The neutrinos free-streamed from that point on and only recently became non-relativistic in the late universe. Next, electrons and neutrinos annihilated (up to one part in a billion, due to the previously produced matter-antimatter asymmetry) when the temperature dropped below the electron mass. Their annihilation reheated (or rather: slowed down the cooling of) the remaining bath of nucleons, electrons and photons to a temperature slightly higher than those of neutrinos. Eventually, the nucleons formed (quasi-)stable bound states in a process called Big Bang nucleosynthesis (BBN). The remaining stable bound states are the ionized nuclei of the early elements hydrogen (hydrogen-1 ^1H as well as deuterium ^2H) and helium (^3He). To some small fraction also the stable lithium isotope ^7Li has been produced. About three minutes after the end of inflation, the formation of the early elements stalled as the universe had cooled so much that it could no longer sustain nuclear interactions. Today's observations of galaxies with low metallicities in which star formation just started allows us to test the predicted relic abundances of these isotopes observationally. The predictions, being about 75% hydrogen, 25% helium and a little ($\mathcal{O}(10^{-10})$) lithium, match the observed mass abundances very well. This success of the ΛCDM model leaves us with the so-called N_{eff} constraint on new physics at the MeV scale, which we will discuss in section 2.3.6.

Up until 60.000 years after BBN, the universe was dominated by relativistic species, commonly referred to as *radiation*. Since the energy density of radiation dilutes faster than that of non-relativistic species, commonly referred to as *matter*, from this point on cold dark and SM matter then dominated the cosmic energy budget. In fig. 2.2 this energy-redshift dependence is plotted. The redshift will be properly defined in the following section, but can intuitively be understood as the relative factor a wavelength of a photon propagating towards us has been stretched since a given point during cosmic expansion.

About 380.000 years after the end of inflation, the universe became transparent to electromagnetic radiation. Whereas it was transparent to GWs throughout all of the previously discussed processes, photons interacted too frequently with the charged constituents of the plasma for us to still observe them today if emitted before this point in time. Only at recombination, electrons and ionized nuclei formed charge-neutral atoms, allowing photons to free-stream until today. The resulting background of photons can be observed today as the CMB. The CMB can be imagined as an almost perfectly smooth, spherical wall blocking our view towards earlier times. Yet, the CMB embodies a rich source of

*ν -decoupling,
 e^+e^- -annihilation
and BBN*

*Radiation-Matter
equality*

Recombination

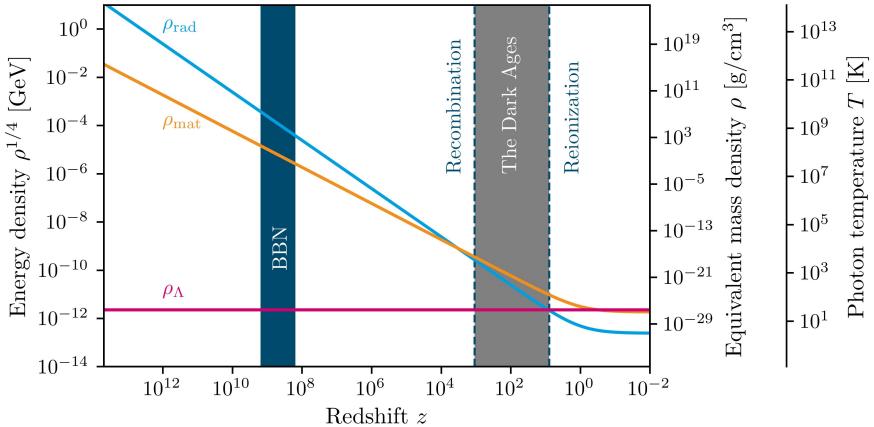


Figure 2.2: Energy density of the primordial relativistic plasma (rad), the matter energy density (mat) and the vacuum energy density (Λ) in Λ CDM. The background module of the CLASS code [41] and cosmological parameters inferred from Planck 2018 data [12] were used to generate this figure.

information about the early universe whose interpretation marked the dawn of precision cosmology: after the accidental detection of the CMB in 1965 by Penzias and Wilson, the COBE satellite confirmed in 1989 that the CMB photons indeed follow an almost perfect black body spectrum with a mean temperature of temperature 2.73 K and anisotropies of only one part in 100.000. Subsequent measurements of the CMB anisotropies with WMAP and eventually the Planck satellite allowed the formulation of the Λ CDM model. In observing the temperature anisotropies and mapping potential cosmic histories to their angular power spectrum, the age and expansion rate of the universe, as well as its energy composition could be inferred (see fig. 2.3).

After the emission of the CMB, the universe remained dark for about 10 to 100 million years. The so-called dark ages ended with the cosmic dawn when baryonic matter fell into the potential well of cold DM, forming the first (“Population-III”) stars which quickly burned hydrogen and helium to form the first 26 elements of the periodic table. At about the same time reionization happened: The first stars emitted ionizing photons, thus *re-ionizing* the surrounding gas clouds. Again, the universe became partially opaque due to the scattering of photons with free electrons. Luckily, the ionization was only partial, such that the CMB is not shielded from us, allowing us to still observe it and its anisotropies today.

Reionization

What follows is the hierarchical structure formation: stars form galaxies, galaxies merged forming bigger galaxies, galaxy clusters, and super-clusters, which are separated by large voids. Eventually, dark energy dominated over matter, slowing down the formation of larger structures. Recently, life started to evolve somewhere in the Milky Way, which learned how to write PhD theses.

Today

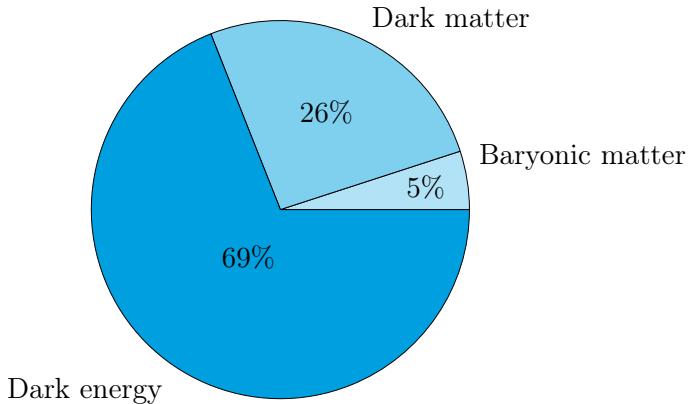


Figure 2.3: Today's energy content of the universe, based on the cosmological parameters inferred from Planck 2018 data [12]. The contribution from radiation is $\mathcal{O}(0.1)\%$ and was hence neglected in this graphic.

2.2 OPEN QUESTIONS

Success and open problems

The Λ CDM model, only being defined through six parameters inferred by observations [12], is the easiest model able to explain almost all cosmological observables so far. In particular, it precisely explains the power-spectrum of CMB anisotropies and the early element abundances produced in BBN. It further successfully describes the distribution of galaxies and large-scale structure in our universe, as well as the observed accelerated expansion of our universe. In conclusion, Λ CDM has been remarkably successful in providing a consistent and comprehensive description of our universe, explaining a wide range of seemingly uncorrelated observations like the structure of the CMB and galaxy clustering. Yet, there are several pieces of evidence that Λ CDM will not remain the cosmological model preferred by data. The most noteworthy tension today is between local measurements of the current expansion velocity (≈ 73 km/s/Mpc) and the inferred Hubble rate today from CMB data (≈ 68 km/s/Mpc). Further, there is a mismatch between the predicted and observed abundances of galaxy clusters. The two tensions are referred to as H_0 and σ_8 tension, respectively. To date there is no consensus on how to reconcile those significant tensions in a sole cosmological model which is statistically preferred over Λ CDM [42]. There is, however, serious hope that data from the James Webb Space Telescope (JWST) will either help in constructing a new cosmic distance ladder which could either resolve the tensions or provide persuasive evidence for new physics [43, 44].

Further, Λ CDM fundamentally relies on the existence of DM, dark energy and an inflaton field, neither of which have been directly observed. The case of DM will be of particular importance for this thesis. It therefore makes sense to briefly review the key pieces of evidence for its existence: Cold DM is not only motivated through its explanatory power when fitting the CMB data to a cosmological model [12], but

further due to several independent sources of evidence: The historically first motivation to introduce this form of non-relativistic (i.e., *cold*) and non-interactive (i.e, *dark*) form of matter was to explain the faster-than-expected movement of galaxies within the Coma cluster by Fritz Zwicky in 1933 [45]. Further evidence for the existence of DM was found by Vera Rubin and collaborators in 1970, when they found that observations of galactic rotation curves contradicted expectations based on visible matter only [46]. Whereas these phenomena alone could also be interpreted as a deviation from the Newtonian law of gravity at large distances, several other phenomena motivate the hypothesized existence of DM. Studies of galaxy clusters show that the majority of the mass inferred from gravitational lensing is not associated with visible matter [47–50]. Further, observations of the bullet cluster provide a striking piece of evidence in favor of DM: Due to the collision of two galaxy clusters, the distribution of baryonic matter within the bullet cluster, visible as hot gas, differs from the distribution of mass as inferred through gravitational lensing [51–53]. Here, we only listed the key pieces of evidence favoring the existence of dark matter. A thorough review of the mentioned phenomena and other sources of evidence can be found for instance in refs. [54, 55].

The evidence for DM

The shortcoming of the SM to provide a viable DM candidate motivates the prediction of new particle species, the arguably most famous one being the WIMP. By now, the WIMP has been subject of many experimental searches, which however only resulted in null-results and exclusion bounds. These, in turn, put the WIMP idea under pressure [56]. Several other DM candidates have been proposed, such as sterile neutrinos, axion-like particles, dark photons, several super-symmetric candidates (such as neutralinos, sneutrinos, gravitinos and axinos), superheavy and composite DM particles, as well as PBHs to just name a few [54, 55].

What is DM made of?

There has been little progress in settling the debate what DM is composed of. This thesis is aimed at contributing to the question how the advent of GW cosmology can hopefully shed light on the still mysterious origin of DM. In the following sections 2.3 and 2.4 we will start by introducing this new sub-discipline of cosmology.

2.3 THE HOMOGENEOUS UNIVERSE

As introduced above, the evolution of our universe can be derived from coupling Einstein's theory of GR with the matter content given by the SM. We will start by solving the Einstein equations for a homogeneous and isotropic perfect fluid and find the Friedmann equations. Those will fix the background evolution and set the stage for the thermodynamic description of the primordial plasma of elementary particles presented in section 2.3.2. In section 2.3.3 we will specifically review which impact dark sector (DS) particles could have had on the primordial plasma. As

Setting the stage for particle cosmology

cosmology became a mature field over the last decades, the derivations this abbreviated introduction to cosmology is based on can be found in countless excellent books, see e.g. [13, 37, 57]. GWs and other metric perturbations only played a negligible role throughout the expansion history, such that their impact can be neglected within this section. We will discuss them in the following section 2.4.

2.3.1 The FLRW metric

According to GR, the gravitational interaction follows the field equations

$$m_{\text{Pl}}^2 G_{\mu\nu} = T_{\mu\nu} \quad (2.1)$$

The Einstein equations

with the Einstein tensor $G_{\mu\nu}$ being a non-linear function of space-time derivatives of the metric $g_{\mu\nu}$ and the energy-momentum tensor $T_{\mu\nu}$ describing the density and flux of energy and momentum through spacetime. As the cosmological constant Λ and the intrinsic curvature of spacetime are irrelevant for the expansion history of the primordial universe, their effect will be ignored in the following discussion.

The FLRW universe

Observations indicate that our universe is spatially homogeneous and isotropic on large scales. Focussing our discussion on these sufficiently large scales, the metric and the energy-momentum tensor are both subject to tight constraints: Due to homogeneity both quantities can only depend on time, but not on a specific position in space. Isotropy further requires them to be diagonal, where the remaining spatial components are identical. The most general metric that fulfills these criteria for the case of a flat universe is the Friedmann–Lemaître–Robertson–Walker (FLRW) metric

$$g_{\mu\nu}(t) = \text{diag}[-1, a(t), a(t), a(t)]. \quad (2.2)$$

The Friedmann equations

The function $a(t)$ is referred to as the scale factor and translates between physical and coordinate distances. It is related to the previously mentioned redshift z of photons from a given cosmic epoch through $a_0/a = 1 + z$, where a_0 is today's scale factor, which is conveniently set to $a_0 = 1$. The energy-momentum tensor is given by

$$T_{\mu\nu}(t) = \text{diag}[-\rho(t), P(t), P(t), P(t)], \quad (2.3)$$

where ρ (P) is the energy density (pressure) of the perfect fluid as which the hot plasma of the early universe is described. Inserting these quantities into the Einstein field eqs. (2.1) yields two independent differential equations for the scale factor, which are referred to as the first and second Friedmann equation. They can be written as

$$H \equiv \frac{\dot{a}}{a} = \sqrt{\frac{\rho}{3m_{\text{Pl}}^2}} \quad \text{and} \quad \dot{\rho} + 3H(\rho + P) = 0. \quad (2.4)$$

The first equation defines the Hubble rate $H(t)$, which works as a measure for the expansion rate of the universe at a given point in time.

Today's Hubble rate is commonly referred to as the Hubble constant H_0 . The above set of equations neatly captures the meaning of Wald's famous quote [58]: "Matter tells space how to curve and space tells matter how to move". Whereas the first Friedmann equation can be identified with the first statement, the second equation maps to the second statement: The energy density of the primordial plasma makes space expand, which in turn leads to the dilution of the plasma.

In the second Friedmann eq. (2.4), not only the energy density, but also the pressure of the primordial plasma enters, making the dimensionless equation of state parameter $w = P/\rho$ the decisive quantity for determining the speed with which any form of energy dilutes. For an equation of state $w \neq -1$, we find that $\rho \propto a^{-3(1+w)}$, $a \propto t^{2(1+w)/3}$ and $H = 2(1+w)/(3t)$ by solving the set of the two Friedmann equations. Of particular importance are the cases $w = \frac{1}{3}$ and $w = 0$, corresponding to a fluid dominated by relativistic and non-relativistic particles, respectively. More conventionally, these fluids are referred to radiation and matter. In these cases the energy density evolves like $\rho_{\text{rad}} \propto a^{-4}$ and $\rho_{\text{mat}} \propto a^{-3}$. The case $w = -1$ corresponds to the case of vacuum energy: If vacuum energy dominates over all other energy sources, as in the present universe, the scale factor increases exponentially with time, as vacuum energy does not get diluted by its own induced Hubble expansion.

The total energy density of our universe today is referred to as $\rho_0 = \rho_{c,0} \equiv 3m_{\text{Pl}}^2 H_0^2 \simeq 10^{-26} \text{ kg/m}^3$, where the Friedmann equation was used to relate its value to the Hubble constant H_0 . If today's mean energy density were larger (smaller) than this critical value, but H_0 remained the same, a positive (negative) intrinsic curvature of Euclidean space would need to be introduced. Measurements of the Planck satellite are however consistent with our universe being intrinsically flat [12].

We further introduce the normalized energy density $\Omega_i = \rho_{i,0}/\rho_{c,0}$, such that $\sum_i \Omega_i = 1$, where the sum runs over all sources ρ_i contributing to the total energy density ρ . The current energy densities in radiation, (dark and visible) matter, and dark energy have been inferred from Planck 2018 data [12] and read $\Omega_{\text{rad}} = 9 \cdot 10^{-5}$, $\Omega_{\text{mat}} = 0.31$, and $\Omega_\Lambda = 0.69$, c.f. fig. 2.3. In this thesis, the Hubble constant $H_0 = 67.7 \text{ km/s/Mpc} = 2.1 \cdot 10^{-42} \text{ GeV}$ (based on TT,TE,EE+lowE+lensing+BAO data) is used [12]. In more conventional units, the inverse Hubble constant reads $H_0^{-1} = 4.4 \text{ Gpc} = 14.5 \text{ Gyr}$, providing us with a rough estimate of the size and age of the observable universe, ignoring the redshift dependence of its energy content. Integrating the Friedmann equation with the actual energy content described by the above values for Ω_{rad} , Ω_{mat} and Ω_Λ leads to the current age of the universe, $t_0 = 13.8 \text{ Gyr}$.

*Equation of state:
a song of fire, ice
and dark energy*

We live in flatland

Today's universe

2.3.2 Thermodynamics of the primordial plasma

Particles: the actors on the FLRW stage

So far we assumed that the energy-momentum tensor on the right-hand side of the Einstein eqs. (2.1) is that of a generic perfect fluid, which is only characterized by a time-dependent energy density and pressure. The goal of this section will be to derive those two expressions for the particle content of a given QFT, in particular of the SM and DS extensions of it.

The primordial plasma has high number densities and typically also high interaction rates between the different particle species. It can hence be described using the tools of statistical mechanics. As such, we are not interested in the dynamics of individual particles' motions but instead how their distribution changes over time. The distribution $f_a(\mathbf{x}, \mathbf{p}, t)$ of positions and momenta of a given particle species a at a time t is defined such that in a given spatial volume \mathcal{V}_x in a three-dimensional momentum range \mathcal{V}_p , the number of particles is defined as

$$N_a(t) = \frac{g_a}{(2\pi)^3} \int_{\mathcal{V}_x} d^3x \int_{\mathcal{V}_p} d^3p f_a(\mathbf{x}, \mathbf{p}, t) \quad (2.5)$$

The Boltzmann equation

with g_a being the internal degrees of freedom of species a . We now impose that the statistical distribution f_a fulfills the Boltzmann equation

$$\mathcal{L}[f_a] = \mathcal{C}[f_a, \dots]. \quad (2.6)$$

The Liouville operator \mathcal{L} on the left-hand side relates the geodesics, given a specific metric tensor $g_{\mu\nu}$, of individual particles of species a to their full distribution f_a . The collision operator \mathcal{C} instead relates the changes in the distribution function f_a with those of all other particle species, indicated by the ellipsis in the above equation. The right-hand side describes particle interactions, i.e. decays and scatterings of a , and can thus be understood as the way QFT enters into our description of the early universe.

For an FLRW metric, in which $f_a(\mathbf{x}, \mathbf{p}, t) = f_a(p, t)$ due to isotropy and homogeneity, the Boltzmann equation reads

$$\partial_t f_a(p, t) - H(t) p \partial_p f_a(p, t) = \frac{\mathcal{C}[f_a, \dots](p, t)}{E_a}, \quad (2.7)$$

where $E_a^2 = m_a^2 + p^2$ is the energy of a single a particle with momentum p . In practice, this form of the Boltzmann equation is still too generic to infer the dynamics of observable quantities from it. What is often done, is to compute moments of the above partial differential equation in order to obtain a set of ordinary differential equations.

Thermodynamic quantities

The most relevant moments of the distribution f_a in thermodynamics are the number density, the energy density, the pressure and the entropy density of a given species a . They can be obtained through

$$n_a(t) = \frac{g_a}{2\pi^2} \int_0^\infty p^2 f_a(p, t) dp, \quad (2.8a)$$

$$\rho_a(t) = \frac{g_a}{2\pi^2} \int_0^\infty p^2 \sqrt{p^2 + m_a^2} f_a(p, t) dp, \quad (2.8b)$$

$$P_a(t) = \frac{1}{3} \frac{g_a}{2\pi^2} \int_0^\infty \frac{p^2}{\sqrt{p^2 + m_a^2}} f_a(p, t) dp \quad \text{and} \quad (2.8c)$$

$$s_a(t) = -\frac{g_a}{2\pi^2} \int_0^\infty p^2 [f_a \log(f_a) \mp (1 \pm f_a) \log(1 \pm f_a)] dp. \quad (2.8d)$$

The upper (lower) signs correspond to a bosonic (fermionic) species a . Integrating the Boltzmann equation in order to obtain equations of motion for the above quantities, one finds

$$\dot{n}_a + 3Hn_a = \frac{g_a}{2\pi^2} \int_0^\infty p^2 \frac{C[f_a, \dots](p, t)}{\sqrt{p^2 + m_a^2}} dp, \quad (2.9a)$$

$$\dot{\rho}_a + 3H(\rho_a + P_a) = \frac{g_a}{2\pi^2} \int_0^\infty p^2 C[f_a, \dots](p, t) dp \quad (2.9b)$$

$$\dot{s}_a + 3Hs_a = -\frac{g_a}{2\pi^2} \int_0^\infty p^2 \frac{C[f_a, \dots](p, t)}{\sqrt{p^2 + m_a^2}} \log\left(\frac{f_a(p, t)}{1 \pm f_a(p, t)}\right) dp. \quad (2.9c)$$

Note that this set of equations for \dot{n}_a , $\dot{\rho}_a$, and \dot{s}_a does not necessarily simplify our way to the original goal of simplifying the full Boltzmann equation. Luckily, often the right-hand side can be approximated or be expressed in terms of the same derived quantities n_a , ρ_a and s_a , allowing an enormous decrease in complexity of the problem. In the case of no particle interactions, $C = 0$, one finds that the comoving particle number density $N_a = n_a a^3$ and the comoving entropy density $S_a = s_a a^3$ are conserved. For a pressure-less non-interactive species, one further finds that the comoving energy density $\rho_a a^3$ is conserved. For radiation instead, for which $P_a = \rho_a/3$, this is not the case. In the latter case, the radiation pressure makes the energy density decrease faster and $\rho_a a^4$ is conserved. The underlying physical process is the redshift of the radiation to less energetic, lower frequencies.

If there is a non-zero collision term for species a , the term $\dot{q}_a \equiv \dot{\rho}_a + 3H(\rho_a + P_a)$ will act as a heat source or sink for other species. The quantity \dot{q}_a is hence often referred to as a heating rate. Comparing the second Friedmann eq. (2.4) with the expression for \dot{q}_a , we can deduce that the total heating rate $\dot{q} = \sum_a \dot{q}_a = 0$ vanishes, such that the total heat $Q = \text{const}$ in the universe is conserved, where $\dot{Q} = \dot{q}a^3$. The Hubble expansion of the universe can hence be understood as an adiabatic expansion in the sense of thermodynamics, slowly cooling down its constituents.

In the case of frequent interactions, the plasma is thermalized. Frequent interactions here refer to an interaction rate which exceeds the Hubble rate, $H \gg \Gamma$. In this case, the distributions can be shown to follow a thermal Bose-Einstein or Fermi-Dirac distribution,

$$f_a(p, t) = \left[\exp\left(\frac{E_a(p) - \mu_a(t)}{T_a(t)}\right) \mp 1 \right]^{-1} \quad (2.10)$$

Moments of the Boltzmann equation

The conservation of particle number, entropy and energy

Adiabatic expansion

Thermal equilibrium

with $\mu_a(t)$ being the chemical potential of a . Depending on the strength of interactions allowed by the underlying QFT, the temperature $T_a(t)$ either is specific to the species a or shared between several species. In case a only has frequent self-interactions, but no sufficiently fast interactions with other species, T_a is specific for a . In case there are frequent interactions involving also another species b , the two species thermalize and obtain a common temperature $T_a = T_b$. For the sake of generality, we will for now assume individual temperatures for all particle species.

Plugging in the thermal distribution (2.10) into eq. (2.8d) we recover the relation

$$\dot{S}_a = a^3(t) \frac{g_a}{2\pi^2} \int_0^\infty p^2 dp \frac{C[f_a, \dots](p, t)}{E_a} \frac{E_a(p) - \mu_a(t)}{T_a(t)} \quad (2.11)$$

$$= \frac{\dot{Q}_a(t) - \mu_a(t)\dot{N}_a(t)}{T_a(t)}, \quad (2.12)$$

where we introduced the comoving heating rate $\dot{Q}_a \equiv \dot{q}_a a^3$ for species a . This is just the second law of thermodynamics, which hence holds for each individual particle species in a comoving volume. From this we can conclude that a particle species' entropy is conserved individually, when there is no heat transfer to other particle species, $\dot{Q}_a = 0$, and if the product $\mu_a \dot{N}_a$ vanishes. The latter can be the case if either the chemical potential $\mu_a = 0$ vanishes or if particle number is conserved, $\dot{N}_a = 0$.

Entropy is conserved individually if $\dot{Q}_a = \dot{N}_a = 0$

Similarly, we can obtain the corresponding relation for the intrinsic thermodynamic quantities,

$$s_a(t) = \frac{\rho_a(t) + P_a(t) - \mu_a(t)n_a(t)}{T_a(t)}, \quad (2.13)$$

when plugging the thermal distribution (2.10) into the definitions of n_a , ρ_a , P_a and s_a in eqs. (2.8). We can use this equation to calculate the entropy density of a given particle species in thermal equilibrium.

Limits for hot and cold species in thermal equilibrium

Under the assumption, that a species is fully relativistic, meaning that its mass and chemical potential are much smaller than its temperature, $T_a \gg \mu_a, m_a$, we obtain

$$n_a = g_a \frac{\zeta(3)}{\pi^2} T_a^3(t) \times \begin{cases} 1 & \text{for bosons} \\ \frac{3}{4} & \text{for fermions} \end{cases}, \quad (2.14a)$$

$$\rho_a = g_a \frac{\pi^2}{30} T_a^4(t) \times \begin{cases} 1 & \text{for bosons} \\ \frac{7}{8} & \text{for fermions} \end{cases}, \quad (2.14b)$$

$$P_a = \frac{\rho_a}{3} \quad \text{and} \quad (2.14c)$$

$$s_a = g_a \frac{2\pi^2}{45} T_a^4(t) \times \begin{cases} 1 & \text{for bosons} \\ \frac{7}{8} & \text{for fermions} \end{cases}. \quad (2.14d)$$

If instead the particle species a is non-relativistic, $T_a \ll m_a$, such that $E_a \simeq m_a + p^2/(2m_a)$, one obtains

$$n_a = g_a \left(\frac{m_a T_a}{2\pi} \right)^{3/2} \exp \left[\frac{\mu_a - m_a}{T_a} \right], \quad (2.15a)$$

$$\rho_a = \left(m_a + \frac{3}{2} T_a \right) n_a \quad \text{and} \quad (2.15b)$$

$$P_a = T_a n_a \ll \rho_a. \quad (2.15c)$$

We recover the well-known Stefan-Boltzmann T^4 scaling of the radiation energy density in eq. (2.14b), as well as the Boltzmann-suppression of a non-relativistic species in eq. (2.15a), together with the ideal gas law in eq. (2.15c). We can further identify eq. (2.15b) with the equipartition theorem $\langle E_a \rangle = m_a + \frac{3}{2} T_a$, also including the rest-mass energy m_a of a single a particle. Eventually, in the limit of a being thermally decoupled, $\dot{q}_a = 0$, we can also retrieve the scaling $\rho_a \propto a^{-4}$ and $\rho_a \propto a^{-3}$ from eq. (2.9b) for a being relativistic or non-relativistic, respectively. An individual, decoupled species hence just feels the same Hubble expansion as the whole system of coupled fluids.

Without any further ado, we can now perform a sum over all particle species to obtain an expression for the total energy density, entropy and pressure of the primordial plasma, where each individual constituent particle species is (at least individually) in equilibrium,

$$\rho = \sum_a \rho_a, \quad s = \sum_a s_a, \quad P = \sum_a P_a. \quad (2.16)$$

The quantities ρ and P are those that appeared in the Friedmann eqs. (2.4). Comparing the expressions derived for a relativistic and non-relativistic species in eqs. (2.14b) and (2.15b), we can see that the expansion will be dominantly driven by relativistic particle species, i.e. by those satisfying $m_a \ll T_a$. Particles, whose mass exceeds the temperature T_a will quickly become exponentially suppressed due to the Boltzmann factor in eq. (2.15a), unless they carry a chemical potential, forbidding them to decay or annihilate to lighter states.

Summing over all species

2.3.3 Dark and visible sectors

So far, we only assumed that all particle species can individually be described by a thermal distribution, with a temperature parameter that can differ between different species. Often, however, two particle species a and b carry a common temperature $T_a = T_b$. This is achieved through sufficiently fast interactions between a and b . We will use this condition as a definition of what we consider a *sector* of particles. In particular, we will distinguish *visible sector* from *dark sector* particles. Visible sector particles are understood to be those which interact frequently enough

A definition for the visible sector

with the SM photon to form a thermal bath with them.² For a given temperature T of the photon, this is the case for all relativistic degrees of freedom of the Standard Model, except for neutrinos below MeV temperatures. We will discuss the case of neutrino decoupling towards the end of this subsection.

Analogously, we define a DS to be a thermal bath with a common temperature T_d between frequently interacting particles, which however *do not* have any sufficiently fast interactions with SM particles. In general, the temperature of the SM photon therefore does not coincide with the DS temperature, $T \neq T_d$. Note that this definition made no statements about DM residing in the DS. In particular it should also be noted that it is possible for a DS to start thermalizing with the SM bath only at a certain point in cosmic history. This scenario will occur in the model discussed in detail in chapter 6. For simplicity, we will still refer to those particles as DS states.

The common temperature T of SM states motivates the definition of so-called *effective relativistic bosonic energy and entropy degrees of freedom (dofs)* (or short and less specific, just *effective dofs*) $g_*(T)$ and $h_*(T)$, which satisfy

$$\rho(T) = g_*(T) \rho_{\text{bos}}(T) \quad \text{and} \quad s(T) = h_*(T) s_{\text{bos}}(T), \quad (2.17)$$

where $\rho_{\text{bos}} = \frac{\pi^2}{30} T^4$ and $s_{\text{bos}} = \frac{2\pi^2}{45} T^3$ are the energy and entropy density of a single massless degree of freedom at temperature T following a Bose-Einstein distribution with vanishing chemical potential.

Counting effective dofs

The full expressions for the effective dofs of a set of species with individual temperatures T_a and intrinsic dofs g_a read [59]

$$g_*(T) = \sum_a g_a \mathcal{G}(z_a) \left(\frac{T_a}{T} \right)^4 \quad \text{and} \quad (2.18a)$$

$$h_*(T) = \sum_a g_a \mathcal{H}(z_a) \left(\frac{T_a}{T} \right)^3, \quad (2.18b)$$

where

$$\mathcal{G}(z_a) = \frac{15}{\pi^4} \int_{z_a}^{\infty} \frac{u_a^2 \sqrt{u_a^2 - z_a^2}}{e^{u_a} \pm 1} du_a \quad \text{and} \quad (2.19a)$$

$$\mathcal{H}(z_a) = \frac{3}{4} \mathcal{G}(z_a) + \frac{15}{4\pi^4} \int_{z_a}^{\infty} \frac{(u_a^2 - z_a^2)^{3/2}}{e^{u_a} \pm 1} du_a \quad (2.19b)$$

with $u_a = \sqrt{m_a^2 + p^2}/T$ and $z_a = m_a/T$. The integrals $\mathcal{G}(z)$ and $\mathcal{H}(z)$ can be evaluated numerically and tabulated for a range of z . For small $z \ll 1$ (i.e., $m \ll T$), both functions evaluate to 1 for bosonic species and to 7/8 for fermionic species, whereas for large $z \gg 1$ (i.e., $m \gg T$),

2 For temperatures above that of the electroweak symmetry breaking, when no gauge boson of electromagnetism was existent yet, the analogous massless gauge boson of the hypercharge $U(1)_Y$ is used in this definition of the visible sector.

both functions approach zero exponentially fast. Approximately, the effective degrees of freedom are hence just the sum of internal degrees of freedom of those particles which satisfy $m_a \ll T$ (up to a factor 7/8 for fermions, see eq. (2.14b)).

Note, however, the factors $(T_a/T)^4$ and $(T_a/T)^3$ in eqs. (2.18). Assuming that all particles either can be thought of as visible sector particles with temperature T or DS particles with a distinct temperature $T_d = \xi T$, we obtain

$$g_*(T) = \sum_{a \text{ visible}} g_a \mathcal{G}(z_a) + \xi^4 \sum_{a \text{ dark}} g_a \mathcal{G}(z_a) \quad (2.20a)$$

$$h_*(T) = \sum_{a \text{ visible}} g_a \mathcal{H}(z_a) + \xi^3 \sum_{a \text{ dark}} g_a \mathcal{H}(z_a) . \quad (2.20b)$$

The case of two secluded thermal baths

If the DS is colder (hotter) than the visible sector by a factor ξ , its contribution to the total energy density will be reduced (enhanced) by a Stefan-Boltzmann factor ξ^4 . Analogously, the total entropy density will obtain a contribution that scales with ξ^3 .

One can make the point that also SM neutrinos constitute a DS according to the above definition. As mentioned in section 2.1, neutrinos decouple from the photon bath at a temperature slightly above an MeV. More quantitatively, this happens when the Hubble rate crosses the electroweak interaction rate $\Gamma_\nu \sim G_F^2 T^5$ with $G_F = 1.2 \cdot 10^{-5} \text{ GeV}^{-2}$ for processes like $\nu_e e^- \rightarrow \nu_e e^-$ or $\nu_e \bar{\nu}_e \rightarrow e^+ e^-$ at $T_{\nu\text{-dec}} = 2 - 3 \text{ MeV}$.³ At this point in time, the neutrinos inherit the temperature of the photon bath, approximately retaining their Fermi-Dirac distribution. As they are also far from being non-relativistic ($m_\nu \ll \text{MeV}$), their energy density decreases with a^{-4} and their temperature T_ν hence decreases with a^{-1} . When the photon temperature drops below the electron mass, electrons and positrons annihilate.⁴ By that time the neutrinos can be considered to be fully decoupled. As electrons and positrons can only annihilate into photons then, $e^+ e^- \rightarrow \gamma\gamma$, the photon bath cools slower than expected by Hubble expansion alone. This process is referred to as reheating, as eventually, the photons will have a slightly higher temperature than neutrinos.

Neutrino decoupling

To compute the temperature of the photon bath compared to that of the decoupled neutrinos, we use that entropy is conserved in the bath of SM particles. Before the electron-positron annihilation, the photon (two polarizations), the electron and positron (each having two helicities) and the three neutrino and three antineutrino families, all share the same

Electron-positron annihilation

³ First, the muon and tau neutrinos decouple at a temperature of 3.1 MeV, followed by the electron neutrinos at 1.9 MeV. An accurate calculation as performed in ref. [60] shows that the decoupling temperature is momentum-dependent. We here cite the temperature at which a neutrino with average momentum given their thermal distribution decouples.

⁴ Being more precise, the long Boltzmann tail of high photon energies leads to the production process $\gamma\gamma \rightarrow e^+ e^-$ being efficient until $T = 0.3 \text{ MeV}$ [37].

temperature $T(a_1)$. Using eq. (2.17), the entropy of the plasma is hence given by

$$s_1 = \frac{2\pi^2}{45} T^3(a_1) \left[2 + \frac{7}{8}(2 + 2 + 2 \cdot 3) \right]. \quad (2.21)$$

After the annihilation, only the two photon helicities and six neutrino dofs are left, where the latter have a to-be-determined temperature $T_\nu(a_2)$ differing from the photon temperature $T(a_2)$. Again using eq. (2.17), one finds

$$s_2 = \frac{2\pi^2}{45} \left[2 \times T^3(a_2) + \frac{7}{8} \times 6 \times T_\nu^3(a_2) \right]. \quad (2.22)$$

Using entropy conservation, $s_1 a_1^3 = s_2 a_2^3$, together with the condition that the neutrino temperature drops like $T_\nu \propto 1/a$ as they are a non-interacting, relativistic species, we find that

$$T_\nu(a_2) = \left(\frac{4}{11} \right)^{1/3} T(a_2). \quad (2.23)$$

Since electrons and protons are (*i*) stable, (*ii*) the lightest SM states that have non-zero masses and (*iii*) the only species still present in the primordial plasma, there are no more states that could change the photon temperature and neutrino through the above process known as entropy injection. The ratio $T_\nu/T = (\frac{4}{11})^{1/3}$ hence remains constant up until the point in the late universe when the neutrinos become non-relativistic.

A first encounter with N_{eff}

The energy density carried by the three neutrino and three anti-neutrino degrees of freedoms at a point in time after the electron-positron annihilation is hence given by

$$\rho_\nu(T) = 3 \times 2 \times \frac{7\pi^2}{8 \cdot 30} T_\nu^4 = N_{\text{eff}}^\nu \times \frac{7\pi^2}{4 \cdot 30} \left(\frac{4}{11} \right)^{4/3} T^4. \quad (2.24)$$

Here, we introduced the effective number of neutrinos N_{eff}^ν , which is used to take into account deviations from the ratio $T_\nu/T = (4/11)^{1/3}$. Assuming the simplified calculation were exact, $N_{\text{eff}}^\nu = 3$ would hold for three SM neutrino families. Taking into account that (*i*) the neutrino decoupling is not instantaneous (see footnote 3) such that also neutrinos can be partially reheated through the electron-positron annihilation, (*ii*) that the neutrino distribution function is not a perfect Fermi-Dirac distribution after their decoupling, (*iii*) that neutrinos can oscillate into each other, (*iv*) that electrons and positrons were not completely ultra-relativistic at neutrino decoupling and (*v*) finite-temperature effects changing the electron and photon dispersion relations in the plasma, one instead finds $N_{\text{eff}}^\nu = N_{\text{eff}}^{\text{SM}} = 3.0440 \pm 0.0002$ [61]. In the following we will often encounter N_{eff} again when deriving cosmological constraints on different expansion histories.

In table 2.1, the mass spectrum of SM states as well as their intrinsic dofs is listed. Summing up all visible sector states, one obtains

Particle a	m_a	g_a	Particle a	m_a	g_a
u, \bar{u}	2.16 MeV	12	e^\pm	0.511 MeV	4
d, \bar{d}	4.70 MeV	12	μ^\pm	105.7 MeV	4
s, \bar{s}	93.5 MeV	12	τ^\pm	1.777 GeV	4
c, \bar{c}	1.27 GeV	12	$\nu_1, \bar{\nu}_1$	< 0.082 eV	2
b, \bar{b}	4.18 GeV	12	$\nu_2, \bar{\nu}_2$	< 0.082 eV	2
t, \bar{t}	172.6 GeV	12	$\nu_3, \bar{\nu}_3$	< 0.082 eV	2
g	0	16	h	125 GeV	1
γ	0	2			
W^\pm	80.4 GeV	6			
Z^0	91.2 GeV	3			

Table 2.1: The particle content of the SM after the EWPT. We quote the masses recommended by the particle data group [39] for quarks, charged leptons, gauge bosons and the Higgs boson. The tightest constraint on the sum of neutrino masses $m_{\text{tot}} < 0.082$ eV is imposed by a combination of BOSS, eBOSS and CMB data [62].

$g_{\text{SM}}(T \gg m_t) = h_{\text{SM}}(T \gg m_t) = 106.75$ at high temperatures. At low temperatures instead, the effect of photon reheating during the electron-positron annihilation is relevant and one obtains $g_{\text{SM}}(T \ll T_{\nu\text{-dec}}) = 2 + \frac{7}{8} \times 6 \times \left(\frac{4}{11}\right)^{4/3} \approx 3.4$ and $h_{\text{SM}}(T \ll T_{\nu\text{-dec}}) = 2 + \frac{7}{8} \times 6 \times \left(\frac{4}{11}\right) \approx 3.9$.

In the left panel of fig. 2.4 the effective degrees of freedom of the SM are plotted in dependence of the photon temperature. In particular, the QCD PT can be nicely identified with the drop in g_{SM} around $T = 150$ MeV. The right panel instead shows the impact on the total effective dofs of a minimalistic hot DS with $\xi = 3$ times higher temperature, containing a scalar with $m_\phi = 10^4$ GeV and a vector boson with $m_{A'} = 10^6$ GeV. Due to the factors ξ^3 and ξ^4 in eq. (2.20), already a minimal extension of the SM can have important consequences for the expansion history if $\xi > 1$. Note, however, that for this figure it was assumed that the DS keeps a temperature ratio of $\xi = 3$ to the SM bath over time and that it is able to dispose of its energy density when the DS would become non-relativistic, such that the approximation of thermal equilibrium with vanishing DS chemical potentials holds at all times.

2.3.4 Going out of equilibrium

Apparently, the usual process of massive particle species becoming more and more Boltzmann-suppressed with decreasing temperatures, up to the point where they become negligible in cosmic history, must have exceptions. Otherwise, the energy budget of our universe would only include radiation and vacuum energy today. Our own existence hence

Effective dofs of the SM

Effective dofs from the dark

The origin of species

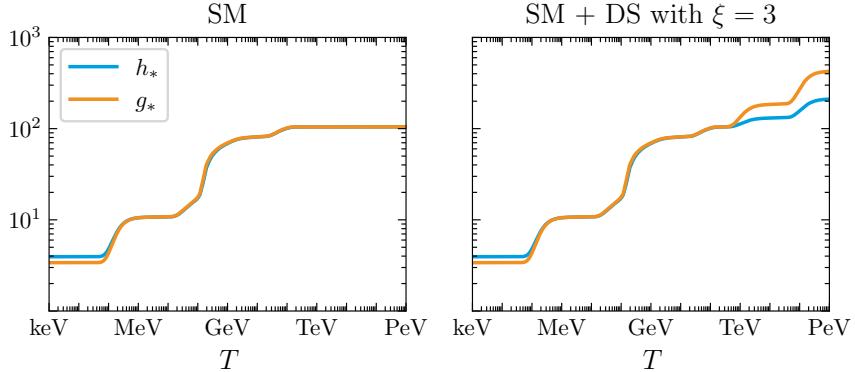


Figure 2.4: Left: Effective dofs of the SM in dependence of temperature. Right: Impact of a hypothetical hot DS containing a dark Higgs boson ($m_\phi = 10 \text{ TeV}$) and a dark photon ($m_A = 1 \text{ PeV}$) with $\xi = T_d/T = 3$ on the effective dofs. To produce this figure, the tabulated data from ref. [59] was used.

leaves us with a proof that a deviation from equilibrium must have occurred. So far, we have not been very specific in our use of the term equilibrium. Let us introduce the following different types of equilibrium in order to be more precise in the following discussion of what it means for a species to *go out of equilibrium*.

KINETIC EQUILIBRIUM refers to a state in which the distribution function $f_a(p, t)$ of a given bosonic (fermionic) species a follows a Bose-Einstein (Fermi-Dirac) distribution (2.10). It is maintained through self-interactions which redistribute the kinetic energy among particles of the species a . Usually, those interactions are elastic scatterings $aa \rightarrow aa$. It implies that the a particles have a common temperature T_a . In an extended sense, one can also speak of two species a and b to be in kinetic equilibrium if they share a common temperature $T_a = T_b$. This equilibrium is typically sustained when elastic scatterings $ab \rightarrow ab$ are efficient. The breakdown of the latter equilibrium is referred to as the (kinetic) decoupling of a and b .

CHEMICAL EQUILIBRIUM is a type of dynamic equilibrium (i.e., all interactions are balanced by their reverse processes, such that the macroscopic properties do not change over time) specific to the particle type. It occurs when there is a balance in particle-type changing interactions, which create and destroy particles of a certain species, such that there is no net change in the number density of that species. Given two particle species a and b as well as their corresponding antiparticles \bar{a} and \bar{b} , chemical equilibrium between a and b is maintained when the interaction rates for $a\bar{a} \leftrightarrow b\bar{b}$ are identical in both directions. For instance, the interaction $e^+e^- \leftrightarrow \gamma\gamma$ sustained the chemical equilibrium between electrons and photons at high temperatures (c.f., footnote 4). More

Decoupling = going out of kinetic equilibrium

Freeze-out = going out of chemical equilibrium

generally, an efficient particle-type changing interaction $ab \leftrightarrow cd$ leads to the species a , b , c and d to be in chemical equilibrium. If such an interaction ceases to be efficient, the corresponding particles are said to “freeze out” since their comoving number densities remain constant in the absence of other interactions.

THERMAL EQUILIBRIUM is a state between two or more particle species (or in the broader thermodynamic sense: arbitrary subsystems of a total system), which do not have any net energy exchange. This implies that there is a common temperature and no net heat flow between the species in thermal equilibrium. It is maintained if both elastic and particle-type changing interaction between the different particle species are efficient.

Each of these terms usually refers to a specific point in time at which the corresponding interactions which maintain the kind of equilibrium are sufficiently fast. What is meant by *staying in equilibrium* is that the particle interaction rate through which the equilibrium is maintained exceeds the Hubble rate over an extended period of time. Vice versa, when this rate drops below the Hubble rate, the corresponding interactions required to maintain the equilibrium can no longer happen sufficiently fast. Depending on the initial type of equilibrium, different phenomena can happen: We have already seen at the example of neutrino decoupling what happens when *kinetic* equilibrium (initially sustained by interactions like $\bar{\nu}e^- \rightarrow \bar{\nu}e^-$) breaks down: there is a decoupling of a species, which will subsequently free-stream. What happens when *chemical* equilibrium breaks down for a species initially in thermal equilibrium with a bath of particles is called the thermal freeze-out⁵ and is the subject of the following subsection.

2.3.5 The freeze-out of dark matter

Starting from an initial state in which all particles were in thermal equilibrium with each other, the thermal freeze-out presents a natural way in which a given particle species can escape its fate of disappearing from the thermal plasma due to its Boltzmann-suppression. What needs to happen can already be inferred from the factor $\exp [(\mu - m)/T]$ in eq. (2.15a): The species needs to develop a non-zero chemical potential which counteracts the Boltzmann-suppression factor $\exp (-m/T)$. For this to occur, chemical equilibrium needs to be broken. In other words, the rate of annihilations of that particle species into particles of the

Chemical potential vs. Boltzmann suppression

⁵ Technically, also the electron-positron annihilation could be referred to as a freeze-out, given that it starts with a departure from chemical equilibrium. However, in this case the annihilation process $e^+e^- \rightarrow \gamma\gamma$ only ceases due to the electron chemical potential required by charge neutrality of the bath of electrons and protons and not due to smallness of couplings as in the case of DM and neutrons.

thermal bath needs to drop below the Hubble rate in order for the species to *freeze out*.

The Boltzmann equation for DM freeze-out

A particularly well-motivated scenario in which freeze-out produces a relic density of particles is the aforementioned WIMP paradigm. Consider a DM fermion χ and its anti-particle $\bar{\chi}$ which can annihilate into an SM fermion ℓ and its associated anti-particle $\bar{\ell}$. For simplicity we will assume that ℓ is essentially massless compared to χ . As ℓ is an SM particle, it will remain in thermal equilibrium throughout the evolution of χ . Hence, $n_\ell = n_\ell^{\text{eq}}$ will hold at all times, where n_ℓ^{eq} refers to the number density in eq. (2.14a). We will further assume that χ and $\bar{\chi}$ have equal abundances $n_{\bar{\chi}} = n_\chi$. To be able to use $T \propto a^{-1}$ we will also assume for simplicity, that there are no other annihilations and entropy injections during freeze-out. It can then be shown, that under these conditions the Boltzmann eq. (2.9a) for the number density reads (see, e.g. [13, 54])

$$\frac{1}{a^3} \frac{d(n_\chi a^3)}{dt} = -\langle \sigma v \rangle \left[n_\chi^2 - (n_\chi^{\text{eq}})^2 \right]. \quad (2.25)$$

Here, $\langle \sigma v \rangle$ is the thermally averaged⁶ cross section of the annihilation process. To make the above equation dimensionless in order to solve it numerically, one usually introduces the yield $Y_\chi = n_\chi/s$ and the time parameter $x = m_\chi/T$. One then obtains the Riccati equation

$$\frac{dY_\chi}{dx} = -\frac{\lambda}{x^2} \left[Y_\chi^2 - (Y_\chi^{\text{eq}})^2 \right]. \quad (2.26)$$

Analytical solutions

Here, we further introduced the dimensionless annihilation-rate-to-Hubble-rate ratio $\lambda = \Gamma_{\text{ann}}(m_\chi)/H(m_\chi)$, where $\Gamma_{\text{ann}}(T) = \langle \sigma v \rangle n_\chi$ and we assumed $\Gamma_{\text{ann}}(T) \simeq \Gamma_{\text{ann}}(m_\chi)$ to be a constant for simplicity. Fig. 2.5 shows numerical solutions to the Riccati equation for a range of interaction rates. One characteristic feature of freeze-out is that it happens at a time $x_f = \mathcal{O}(10 - 20)$, i.e. when the number density of χ is suppressed by $\exp(-x_f)$. Moreover, one can see that the yield produced via freeze-out decreases proportionally to λ^{-1} . This is due to a later freeze-out in case of a stronger coupling between χ and ℓ : The longer χ stays in chemical equilibrium with ℓ , the longer its number density can get Boltzmann-suppressed. Ignoring the second term in eq. (2.26), one can indeed obtain the analytical estimate $Y_\chi^\infty \simeq x_f/\lambda$ for the relic yield of χ .

Using this analytical approximation, one can convert the relic yield Y_χ^∞ to a relic abundance Ω_χ through [54]

$$\Omega_\chi = \frac{m_\chi}{3m_{\text{Pl}}^2 H_0^2} Y_\chi^\infty T_0^3 \frac{h_*(T_0)}{h_*(T_f)} \sim 0.1 \frac{x_f}{\sqrt{g_*(m_\chi)}} \frac{10^{-8} \text{ GeV}^{-2}}{\langle \sigma v \rangle}. \quad (2.27)$$

To arrive at this approximate expression for the contribution of χ to today's matter abundance, we assumed that entropy is conserved

⁶ The averaging refers to an average over relative velocities between χ and $\bar{\chi}$, on which the cross section $\sigma(v)$ will depend. Given that χ is in kinetic equilibrium at all times, the distribution of relative velocities can be inferred from the thermal distributions (2.10).

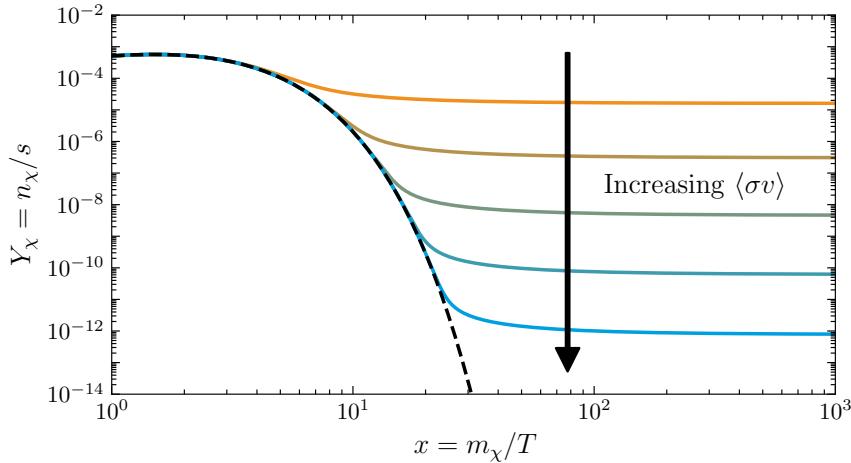


Figure 2.5: Freeze-out of a scalar χ with $m_\chi = 1 \text{ TeV}$ for five thermally averaged cross sections $\langle\sigma v\rangle = 10^{-17} - 10^{-9} \text{ GeV}^{-2}$ (blue to red curves), each being a factor 10^3 higher than the previous one. The dashed black line depicts $Y_\chi^{\text{eq}} = n_\chi^{\text{eq}}/s$, which χ would follow if it stayed in chemical equilibrium with the SM bath.

between the freeze-out and today, used that $g_*(m_\chi) \simeq h_*(m_\chi)$ as well as $h_*(T_0) = 3.9$, and plugged in the CMB temperature $T_0 = 2.7 \text{ K}$ and the Hubble constant $H_0 = 67.7 \text{ km/s/Mpc}$ [12].

From this analytical estimate we can see that the observed DM abundance $\Omega_{\text{DM}} h^2 = 0.12$ can be produced by a weakly interacting particle with annihilation rates into the SM bath of $\sqrt{\langle\sigma v\rangle} \sim 10^{-4} \text{ GeV}^{-1} \sim 0.1 G_F^{1/2}$. This unexpected coincidence of producing the observed DM relic density by a weakly interacting massive particle is usually referred to as the *WIMP miracle*.

The thermal freeze-out mechanism can be seen as the starting point of a whole variety of more involved DM production mechanisms. An easy extension of the idea is based on the assumption that DM never was in thermal equilibrium with the SM bath in the first place. This idea led to the discovery of the freeze-in mechanism, which is relevant for DM candidates with even lower couplings to the SM [63, 64]. Other noteworthy variations of the WIMP idea are for instance based on the assumption of the DM mass being lower than that of any SM state it can annihilate into (i.e., forbidden DM [65, 66]), number-changing processes in the DS (i.e., cannibal DM [67, 68] and pandemic DM [69]) or a non- Λ CDM expansion history after DM production (i.e., homeopathic DM [70, 71]). An extension of the WIMP idea that will be of particular importance for the course of this thesis is the freeze-out of a WIMP from a DS bath instead of the SM bath. This idea will be discussed in chapter 6.

The WIMP miracle

Variations of the WIMP idea

2.3.6 The Big Bang nucleosynthesis

*Neutron freeze-out
sets the initial
condition for BBN*

*The freeze-out of
neutron-proton
conversions*

*The dependence of
the neutron
freeze-out on N_{eff}*

In fact, the freeze-out mechanism is not only relevant for the hypothesized production of WIMP DM, but also plays a central role within BBN. Around the MeV temperature scale, the freeze-out of the processes $n\nu_e \leftrightarrow p^+e^-$ and $ne^+ \leftrightarrow p^+\bar{\nu}_e$ sets the initial abundance of neutrons, which directly influences the early element abundance produced in BBN.

The chemical equilibrium between protons and neutrons is initially maintained as long as the rate $\Gamma_{pe \rightarrow n\nu}$ is much larger than the Hubble rate $H(t)$. Ignoring the chemical potentials of electrons and neutrinos, the chemical equilibrium of neutrons and protons implies $\mu_n = \mu_p$. Using eq. (2.15a), the ratio of their number densities hence follows

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p} \right)^{3/2} e^{-(m_n - m_p)/T} \simeq e^{-Q/T} \quad (2.28)$$

with the mass difference $Q \equiv m_n - m_p = 1.30 \text{ MeV}$. For temperatures above an MeV, neutrons and protons hence have an equivalent number density, whereas for lower temperatures the relative neutron fraction decreases exponentially. If the interactions converting neutrons to protons were infinitely quick, the chemical equilibrium would be maintained and the neutron abundance would have basically vanished before the onset of BBN starting with the production of deuterium at $T \lesssim 0.1 \text{ MeV}$, as we will discuss in more detail below. What happens instead is that the conversion reaction freezes out at $T_f \simeq 0.8 \text{ MeV}$, leaving us with $n_n/n_p \simeq 0.20$, i.e. roughly five times more protons than neutrons at that temperature [13].

A quick estimate of the neutron freeze-out temperature T_f can be obtained by comparing the rate $\Gamma_{\text{weak}} \propto G_F^2 T^5$ for the weak processes $n\nu_e \rightarrow p^+e^-$ and $ne^+ \rightarrow p^+\bar{\nu}_e$ with the Hubble rate $H(T)$. Next to the energy density $\rho_\gamma + \rho_\nu$ from photons and neutrinos dominating the cosmic expansion velocity, we also consider an additional energy density ρ_{extra} contributing to the Hubble parameter through the Friedmann eq. (2.4). In eq. (2.24) we introduced the parameter N_{eff}^ν to quantify the impact of the number of neutrino species on the temperature ratio between neutrinos and photons after electron-positron annihilation. We now want to parameterize the amount of additional energy density ρ_{extra} using the analogous parameter N_{eff} defined through

$$\rho_\nu + \rho_{\text{extra}} \equiv N_{\text{eff}} \times \frac{7}{8} \left(\frac{4}{11} \right)^{4/3} \rho_\gamma, \quad (2.29)$$

such that the extra energy can be expressed as⁷

$$\rho_{\text{extra}} = \Delta N_{\text{eff}} \times \frac{7}{8} \left(\frac{4}{11} \right)^{4/3} \rho_\gamma \quad \text{where} \quad \Delta N_{\text{eff}} \equiv N_{\text{eff}} - N_{\text{eff}}^{\text{SM}}. \quad (2.30)$$

⁷ Note that in general ΔN_{eff} is a function of temperature for ρ_{extra} having an arbitrary temperature dependence. Only for ρ_{extra} redshifting like radiation ΔN_{eff} is a constant.

Plugging in $\rho = \rho_\gamma + \rho_\nu + \rho_{\text{extra}}$ into the first Friedmann eq. (2.4) to get an expression for $H(T)$ and requiring freeze-out to happen when $\Gamma(T_f) \simeq H(T_f)$ yields

$$T_f \simeq \left(\frac{\pi^2}{45} \right)^{1/6} \left[1 + \frac{7}{8} \left(\frac{4}{11} \right)^{4/3} N_{\text{eff}} \right]^{1/6} \frac{1}{(G_F^2 m_{\text{Pl}})^{1/3}}. \quad (2.31)$$

For an increase in N_{eff} with respect to $N_{\text{eff}}^{\text{SM}}$, corresponding to a positive contribution ρ_{extra} and $\Delta N_{\text{eff}} > 0$, the Hubble expansion is faster, such that the neutron freeze-out occurs at a slightly higher temperature. Due to the exponential sensitivity of the neutron-to-proton ration in eq. (2.28), N_{eff} is hence an important input parameter in any prediction of the early element abundances. In a more precise calculation the $\mathcal{O}(1)$ pre-factor and the full temperature dependence of the weak interaction rate, which we here just approximated as $\Gamma_{\text{weak}} \propto G_F^2 T^5$, can be determined analytically. Our simplified treatment yields a freeze-out temperature of 1.2 MeV for $N_{\text{eff}} = N_{\text{eff}}^{\text{SM}}$, which approximately agrees with the previously stated temperature $T_f = 0.8$ MeV which results from actually solving the Boltzmann equation, see for instance ref. [13].

A faster expansion implies an earlier freeze-out

The onset of BBN is marked by the formation of deuterium nuclei, composed of one proton and one neutron, as the fusion of two neutrons only leads to a very unstable bound state which immediately decays. Two-proton fusion is also inefficient, as the Coulomb barrier would need to be overcome. Hence, only as soon as deuterium nuclei have formed, more massive elements can form. The deuterium synthesis through the interaction $n + p^+ \rightarrow D^+ + \gamma$ is delayed, however, until a temperature of around $T_D \simeq 0.1$ MeV due to the sheer abundance of photons in the plasma. The baryon-to-photon ratio $\eta \equiv n_\gamma/n_b$ was determined by CMB measurements to be of order $\eta \sim 10^{-9}$ [12], meaning that there are roughly a billion times more photons than neutrons and protons present in the plasma. Due to the long Boltzmann-tail of the photon distribution at high photon energies, the deuterium photo-disintegration only freezes out at a temperature T_D much lower than the deuterium binding energy $B_D \equiv m_n + m_p - m_B \simeq 2.22$ MeV. As all other nuclear processes rely on the presence of deuterium due to the above reasoning, this phenomenon became known as the deuterium bottleneck.

The onset of BBN: the deuterium bottleneck

Eventually, since the binding energy of helium is larger than that of deuterium, the production rate of helium exceeds that of deuterium. Virtually all neutrons present at the onset of nucleosynthesis will hence end up in ${}^4\text{He}$. The helium mass fraction

$$\mathcal{Y}_p \equiv \frac{4n_{\text{He}}}{n_n + n_p} = \frac{2n_n}{n_n + n_p} \simeq 2 \frac{n_n/n_p}{1 + n_n/n_p} \Big|_D \quad (2.32)$$

therefore directly connects the observable relative abundance of ${}^4\text{He}$ to the neutron-to-proton ratio at the time of the onset of BBN, indicated by a D for the deuterium bottleneck in the above equation. Due to the finite lifetime $\tau_n = 887$ s of the neutron, the neutron-to-proton

Virtually all neutrons end up in ${}^4\text{He}$

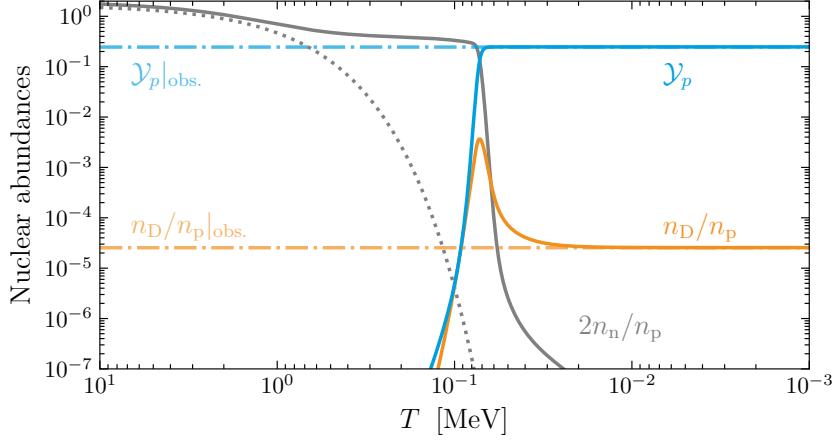


Figure 2.6: Evolution of the neutron-to-proton ratio (gray), the helium mass fraction (blue), and the deuterium fraction (orange). The equilibrium neutron-to-proton ratio, cf. eq. (2.28), is plotted as a dotted gray line. After the neutron freeze-out at $T_f \simeq 0.8$ MeV, n_n/n_p only decreases slightly due to the neutron decay up to the point when the deuterium photo-disintegration becomes inefficient at $T_D \simeq 0.1$ MeV. After this point, deuterium can form efficiently and fuse to helium. The data for producing this plot was kindly provided by Frederik Depta.

ratio slightly drops between the neutron freeze-out at $T_f \simeq 0.8$ MeV and the formation of deuterium at $T_D \simeq 0.1$ MeV from $n_n/n_p \simeq 0.20$ to $n_n/n_p \simeq 0.14$. Plugging this into the above equation for the helium mass fraction, we obtain $\mathcal{Y}_p = 0.24$, which coincides with the observed value $\mathcal{Y}_p = 0.245 \pm 0.003$. Analogously, one can further set up and solve the Boltzmann equation for the production of heavier elements like deuterium, lithium and beryllium. This task is usually done by tools like **Primat** [72] or **AlterBBN** [73], which solve the tower of these coupled Boltzmann equations including many hundred nuclear interactions.

Precision cosmology

In fig. 2.6, the evolution of the neutron-to-proton ratio (in gray), the helium mass fraction (in blue) and the deuterium fraction (in orange) are depicted. Note how the observed nuclear abundances precisely match their predictions. This success of BBN to predict nuclear abundances (shown as horizontal dash-dotted lines) up to the percent-level is often referred to by the term precision cosmology.

Above, we showed that any form of additional energy ρ_{extra} results in a larger neutron-to-proton ratio at the time of neutron freeze-out. Due to the above reasoning, this results in an overproduction of ${}^4\text{He}$. An increase of N_{eff} by 1 unit (corresponding to the equivalent of one additional neutrino-like species contributing as ρ_{extra}) would increase the Hubble parameter at that time by 7%, leading to a 2% increase in the freeze-out temperature, a 2% increase in the neutron-to-proton ratio and correspondingly also a 2% increase in \mathcal{Y}_p . Since the helium mass

fraction was measured on a percent-level, precise bounds on N_{eff} can be inferred. In our above treatment we have seen, however, that not only N_{eff} , but also the baryon-to-photon ratio η is a crucial ingredient in order to determine a prediction for the early element abundances. In a more careful analysis, see [13], one finds that there is a logarithmic dependence of the predicted helium mass fraction on η , resulting in a degeneracy between N_{eff} and η . This degeneracy can however be resolved when taking into account that the helium fraction is not the only BBN observable. In particular, there is a strong dependence of the deuterium fraction D/H on η , leading to a complementary measurement, solving the degeneracy.

Inferring N_{eff} and η from BBN

Comparing the observations of the early element abundances with their predicted counterparts, one can hence obtain a probability distribution $p(\eta, N_{\text{eff}})$ which depends on both η and N_{eff} . Observational limits on N_{eff} can then be obtained by combining the distribution $p(\eta, N_{\text{eff}})$ inferred from BBN with its counterpart inferred from CMB measurements and marginalizing over η . Eventually, one obtains the observational bound [74]

$$N_{\text{eff}} = 2.941 \pm 0.143 \quad (2.33)$$

at 95% C.L. We will use this limit on N_{eff} in two ways in this thesis: The first use will be in constraining the amplitude of cosmological GWBs in the following section by setting $\rho_{\text{extra}} = \rho_{\text{gw}}$ (cf. eq. (2.82)). In chapter 5 we will instead interpret ρ_{extra} as coming from a DS, whose energy is dominated by a plasma reheated in a PT.

2.4 COSMOLOGICAL GRAVITATIONAL WAVE BACKGROUNDS

In the previous section 2.3 we discussed the expansion history of our universe. We found that it only depends on the sum of all energy densities and pressure contributions of its individual sub-components, which we described through a perfect, hence spatially homogeneous and isotropic fluid. Obviously, however, the present universe is neither homogeneous nor isotropic—instead its matter distribution is clustered into celestial bodies, galaxies and more extended objects, and large voids. Only when describing large enough scales, the energy distribution and hence the right-hand side of the Einstein eqs. (2.1) can be regarded as being independent of a specific position in space. This section will now try to bridge this gap by considering classes of solutions of the Einstein equation which are not based on the assumptions of statistical homogeneity and isotropy. Instead, we will in particular discuss how gravitational waves (GWs) arise when considering deviations from homogeneity and isotropy.

The limits of a perfect fluid description

*A short history of
doubt*

Historically, the prediction that GR features propagating dofs is almost as old as the theory itself. The derivation was first performed by Albert Einstein himself in 1918 [6, 75]. The existence of GWs was, however, facing serious doubts for a long time: Sir Arthur Eddington's statement that GWs would propagate “at the speed of thought” [76], due to some of the wave solutions found by Einstein being gauge artifacts quickly became famous. Together with Nathan Rosen, again even Einstein himself wrote an article in 1936 concluding that GWs could not exist [77]. The question of the correct gauge choice was only solved two decades later by Felix Pirani in 1956 [78], who showed that passing GWs could actually move test masses. This idea, eventually led to the “sticky bead argument” by Richard Feynman at the first GR conference in 1957, which was popularized by Hermann Bondi [79] and convinced physicists at the time that gravitational radiation is indeed a detectable, physical phenomenon. After some erroneous claims for the observation of GWs using so-called Weber bars in 1969 [80], the first indirect evidence for a GW-induced phenomenon came from astrophysics: In 1982, the Hulse-Taylor binary pulsar B1913+16 was shown to have a cumulative shift of its periastron time which precisely followed the prediction of an orbital decay through GW emission [81]. Eventually, the first direct detection of GWs at the Laser Interferometer Gravitational-Wave Observatory (LIGO) in 2014 [5] was made possible through the progress of laser interferometry based on the work of Forward and Weiss in the early 1970s [82, 83].

*Outline of this
chapter*

In the following four subsections we introduce the concept of gravitational radiation with increasing levels of diligence: In section 2.4.1 we review the conceptually easiest and most well-known scenario of GWs being the propagating metric perturbations of the linearized Einstein equations. In order to understand the sourcing mechanism of GWs better we then go over to the more advanced notion of a GW as being a gauge-invariant tensor perturbation after performing a helicity decomposition of the Einstein equations into scalar, vector and tensor perturbations around a flat Minkowski metric. This allows us to identify anisotropic stress to be the source of GWs in section 2.4.2. Even after adding this level of complication, there is yet no notion of the energy density of a GW. Only after allowing the background metric to back-react on the GW by treating it as a dynamical object itself in section 2.4.3, it becomes clear how the energy-momentum tensor of a GW can be computed. In section 2.4.4 we finally consider the case that the background through which GWs propagate is the FLRW metric. This brings us straight to the equations of motion of gravitational radiation in the context of cosmology in section 2.4.5. In section 2.4.6 we then discuss why any GWB from the early cosmos can only be detected as a stochastic signal today. Subsections 2.4.7 and 2.4.8 deal with the different ways a GW spectrum can be described and how a cosmic GWB gets redshifted until today. Eventually, this chapter concludes with a short review of the bounds on stochastic GWBs in section 2.4.9. The main references the following subsections are based on are [11, 14, 37].

2.4.1 Gravitational waves in vacuum

Let us start with the simplest scenario in which gravitational radiation can be found as a solution of the Einstein eqs. (2.1). To start with, note that the field equations are by construction invariant under coordinate transformations $x^\mu \rightarrow x'^\mu(x)$ with x'^μ being an arbitrary smooth function of x^μ . To be more precise, $x'^\mu(x)$ can be an arbitrary diffeomorphism, i.e. it has to be invertible, differentiable and its inverse has to be differentiable. Sometimes, this invariance under coordinate transformations is referred to as the gauge symmetry of GR. Under these transformations, the metric tensor transforms as

$$g_{\mu\nu}(x) \rightarrow g'_{\mu\nu}(x') = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}(x). \quad (2.34)$$

We now consider a small perturbation $h_{\mu\nu}$ around a flat background, $g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x)$, and a coordinate-dependent shift of the reference frame, $x^\mu \rightarrow x'^\mu = x^\mu + \xi^\mu(x)$. In order to speak of small perturbations we further need to require that in our frame $|h_{\mu\nu}(x)| \ll 1$ holds in a sufficiently large region of space-time. After performing this change of coordinates the metric tensor perturbation reads

$$h_{\mu\nu}(x) \rightarrow h'_{\mu\nu}(x') = h_{\mu\nu}(x) - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu) \quad (2.35)$$

to linear order in $h_{\mu\nu}$. The perturbativity condition $|h_{\mu\nu}| \ll 1$ is conserved in this new coordinate frame if $\mathcal{O}(|\partial_\mu \xi_\nu|) \leq \mathcal{O}(|h_{\mu\nu}|)$. In this case we can refer to the specific performed change of coordinates as a symmetry of the whole theory. The underlying theory of linearized gravity is characterized by the Ricci tensor $R_{\mu\nu}$ not only being covariant under general coordinate changes $g_{\mu\nu}(x) \rightarrow g'_{\mu\nu}(x')$ but instead being invariant under $h_{\mu\nu}(x) \rightarrow h'_{\mu\nu}(x')$ up to linear order in $h_{\mu\nu}$.

In linearized gravity, any solution $h_{\mu\nu}$ of the Einstein equations is therefore equivalent to a solution $h'_{\mu\nu}$ in another frame. It is customary to introduce $h \equiv \eta^{\mu\nu} h_{\mu\nu}$ and $\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h$ to express the linearized field equations as

$$\square \bar{h}_{\mu\nu} + \eta_{\mu\nu} \partial^\alpha \partial^\beta \bar{h}_{\alpha\beta} - \partial^\alpha \partial_\nu \bar{h}_{\mu\alpha} - \partial^\alpha \partial_\mu \bar{h}_{\nu\alpha} = -2 \frac{T_{\mu\nu}}{m_{\text{Pl}}^2} \quad (2.36)$$

with the flat-space d'Alembertian $\square = \partial_\mu \partial^\mu$. Note that so far we have not specified the coordinate system we want to work in. In order to simplify our calculations we now change this and will work in a frame which satisfies the Lorentz⁸ gauge $\partial^\nu \bar{h}_{\mu\nu} = 0$. In this coordinate system the field equations simplify to the wave equation

$$\square \bar{h}_{\mu\nu} = -2 \frac{T_{\mu\nu}}{m_{\text{Pl}}^2}. \quad (2.37)$$

The gauge symmetry of GR

Linearizing GR

The linearized field equations

The “Lorentz” gauge

8 Amusingly, this gauge was neither introduced by the Danish physicist Ludvig Valentin Lorenz known for the related condition $\partial_\mu A^\mu = 0$ in electromagnetism, nor the Dutch physicist Hendrik Antoon Lorentz who was still a child when Lorenz first proposed his gauge. Instead it was De Sitter who suggested the gauge to Einstein [84]. Alternative names of this particular and the more general condition $\partial_\mu (g^{\mu\nu} \sqrt{g}) = 0$ are also harmonic gauge, Hilbert gauge and De Donder gauge. Nonetheless, we will use the commonly accepted, though technically incorrect, term “Lorentz gauge” in this thesis.

In order to check whether this choice of coordinate system can be used under the previous specifications, we need to check under which conditions the transformed metric perturbations

$$\bar{h}_{\mu\nu} \rightarrow \bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial_\alpha \xi^\alpha) \quad (2.38)$$

Gauge fixing reduces the dofs to $6 = 10 - 4$

indeed satisfy the gauge condition. We find that the gauge condition $\partial^\nu \bar{h}_{\mu\nu} \stackrel{!}{=} 0$ in the new frame implies $(\partial^\nu \bar{h}_{\mu\nu})' = \partial^\nu \bar{h}_{\mu\nu} - \square \xi_\mu \stackrel{!}{=} 0$. The condition $\square \xi_\mu = \partial^\nu \bar{h}_{\mu\nu}$ for the last equation to hold can be satisfied by fixing the coordinate transformation to

$$\xi_\mu(y) = \int d^4x G(x-y) \partial^\nu \bar{h}_{\mu\nu}(y), \quad (2.39)$$

where $G(x-y)$ is the Green's function of the d'Alembertian, i.e. a solution of $\square_x G(x-y) = \delta^4(x-y)$. Since the gauge condition imposes four distinct constraints, corresponding to the four independent components of $\xi_\mu(y)$ in eq. (2.39), the independent dofs of the symmetric tensor field $\bar{h}_{\mu\nu}$ reduce to $10 - 4 = 6$.

An analogy to electromagnetic waves

Before we proceed with solving the wave equation, note that the previous derivation is in fact completely analogous to what is commonly done for finding a wave equation for the vector field A_μ in electromagnetism: The equations of motion in that case read $\partial_\mu (\partial^\mu A^\nu - \partial^\nu A^\mu) = j^\nu$, which simplifies to $\square A^\mu = j^\mu$ after imposing the Lorentz gauge $\partial_\mu A^\mu = 0$. This gauge leaves the residual gauge freedom $A_\mu \rightarrow A_\mu - \partial_\mu \theta$ untouched, where θ only needs to satisfy $\square \theta = 0$. In vacuum, one finds $\square A^\mu = 0$, such that the residual gauge dof following $\square \theta = 0$ can be used to fix $A^0 = 0$. This makes the gauge condition effectively a transversality condition $\partial_i A^i = 0$, indicating the vanishing mass of the underlying photon field. If instead $j^0 \neq 0$ (i.e., in the presence of charges), one finds that $\square A^0 \neq 0$. In that case A^0 cannot be set to zero using a function θ that satisfies $\square \theta = 0$, complicating the solution of the wave equation.

In vacuum, only $2 = 6 - 4$ dofs remain

Analogously and to simplify the situation for now, we will therefore only consider the case of a vacuum solution of linearized gravity by setting $\square \bar{h}_{\mu\nu} = 0$. This wave equation already tells us that in GR the speed of gravitational radiation is the speed of light. Completely analogous to the above comment on the case of gauge fixing in electromagnetism, note that the Lorentz gauge condition is not spoiled by further constraining the coordinate shift to satisfy $\square \xi_\mu = 0$. Note that under this additional condition, the term $\xi_{\mu\nu} \equiv \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial_\alpha \xi^\alpha$ in brackets in eq. (2.38) satisfies $\square \xi_{\mu\nu} = 0$. We can hence reduce the number of independent components of the metric perturbation $\bar{h}_{\mu\nu}$ yet another time by four conditions, such that only $2 = 6 - 4$ independent components remain.

In particular, we can choose the function ξ^0 such that the trace $\bar{h} = 0$ vanishes, such that $\bar{h}_{\mu\nu} = h_{\mu\nu}$, and choose ξ^i such that $h^{i0} = 0$. The 0-component of the Lorentz gauge condition $\partial^\nu \bar{h}_{\mu\nu} = 0$ then implies $h_{00} = 0$, which means that h_{00} is a constant, being related to a stationary gravitational potential. In this subsection we are only interested in the

propagating solutions of the Einstein equations and will treat the effect of stationary components in $h_{\mu\nu}$ only in the following section. For the moment we can therefore set $h_{00} = 0$. In summary, we chose

$$h^{0\mu} = 0, \quad h^i_i = 0, \quad \partial^i h_{ij} = 0. \quad (2.40)$$

The only contributions to $h_{\mu\nu}$ which do not vanish under this very convenient choice of coordinates are the spatial parts h_{ij} . Since they satisfy the transversality condition $\partial^i h_{ij} = 0$ and the tracelessness condition $h = \eta_{\mu\nu} h^{\mu\nu} = h^i_i = 0$, the above set of conditions is known as the transverse-traceless (TT) gauge. Note that, similar to the above discussion for electromagnetism, the TT gauge cannot be imposed if not in vacuum.

In the TT gauge, the wave equation reads $\square h_{\mu\nu}^{\text{TT}} = 0$ where $h_{\mu\nu}^{\text{TT}}$ is a symmetric tensor field which satisfies the conditions in eq. (2.40). We now want to solve this equation in order to get a feeling for the effect of a GW on test masses. The first TT-condition leaves us with only the spatial parts h_{ij}^{TT} as dynamical fields. For simplicity, we will assume that the GW under considerations is a plane wave $h_{ij}^{\text{TT}}(x) = \text{Re}[e_{ij}(\mathbf{k})e^{ikx}]$ moving in \mathbf{k} -direction. The wave equation requires a dispersion relation $\omega \equiv k_0 = |\mathbf{k}|$, such that $k_\mu = (\omega, \mathbf{k})$. We further assume without loss of generality that the wave propagates in z -direction, such that $\mathbf{k} = (0, 0, \omega)$. The transversality condition $\partial^i h_{ij}^{\text{TT}} = 0$ hence implies $k^i h_{ij}^{\text{TT}} = 0$, such that $h_{3\mu}^{\text{TT}} = h_{\mu 3}^{\text{TT}} = 0$. The only remaining non-zero components h_{ij}^{TT} with $i, j \in \{1, 2\}$ further need to satisfy the tracelessness condition $h_{11}^{\text{TT}} + h_{22}^{\text{TT}} = 0$. We can thus write the solution of the wave equation as

$$h_{\mu\nu}^{\text{TT}}(t, z) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos[\omega(t - z)] \quad (2.41)$$

with the two dofs of a GW described by a $+$ and a \times polarization. The corresponding infinitesimal line element reads

$$\begin{aligned} ds^2 = & dt^2 - dx^2 \{1 + h_+ \cos[\omega(t - z)]\} \\ & - dy^2 \{1 - h_+ \cos[\omega(t - z)]\} \\ & - 2 dx dy h_\times \cos[\omega(t - z)] - dz^2. \end{aligned} \quad (2.42)$$

Now consider two test masses at $(t, x_1, 0, 0)$ and $(t, x_2, 0, 0)$. Their *coordinate distance* in x -direction $L_x = x_2 - x_1$ is constant in time, whereas their proper distance reads $s = L_x \sqrt{1 + h_+ \cos(\omega t)}$. Keeping only the linear order in h_+ , their relative *proper distance* in x -direction hence changes like $\delta s_x = \frac{1}{2} h_+ \cos(\omega t)$. If the two test masses were mirrors reflecting a light beam hence and forth, this proper distance would correspond to an oscillating time needed for the light to reach the other mirror. This is precisely the working principle of GW interferometers. Analogously, in a PTA one test mass is identified with the barycenter of

The TT gauge

The plane wave solution

The effect of a GW on test masses

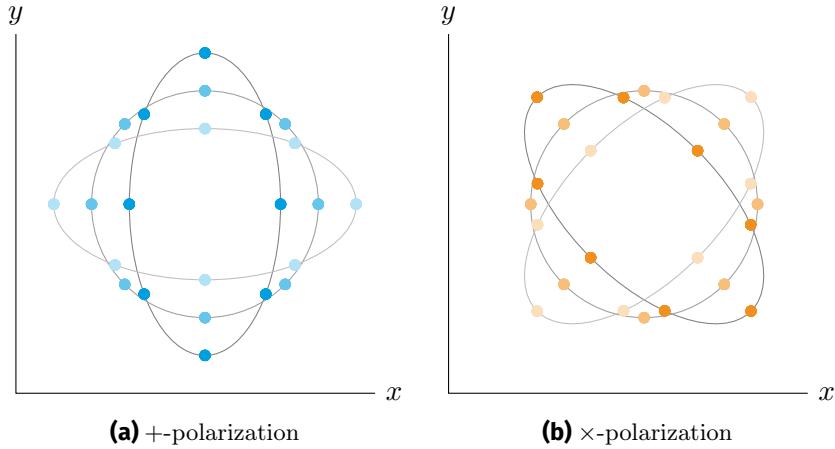


Figure 2.7: Illustration of the effect of a purely $+$ -polarized (left) and \times -polarized GW propagating in z -direction on a ring of test masses lying in the x - y -plane. Time dependence is indicated by the colors' saturation.

our solar system, whereas the observed pulsars correspond to a set of other test masses.

To understand the different effects of the $+$ and \times polarization further, now consider a test mass at the origin $(t, 0, 0, 0)$ and some test masses i each sitting at a set of coordinates $(t, x_i, y_i, 0)$. A GW that only carries a $+$ polarization causes the shifts

$$\delta s_{x,i} \simeq \frac{h_+}{2} x_i \cos(\omega t) \quad \text{and} \quad \delta s_{y,i} \simeq -\frac{h_+}{2} y_i \cos(\omega t) \quad (2.43)$$

in the proper distance. Analogously, the shift in proper distance induced through a \times -polarized GW reads

$$\delta s_{x,i} \simeq \frac{h_\times}{2} y_i \cos(\omega t) \quad \text{and} \quad \delta s_{y,i} \simeq \frac{h_\times}{2} x_i \cos(\omega t). \quad (2.44)$$

A sketch of these effects on a ring of test masses can be found in fig. 2.7.

2.4.2 Helicity decomposition of metric perturbations

In the previous section we found that linearized GR features a wave equation, which we then solved in the absence of sources and other energy distributions. The latter choice offered the tremendous advantage of allowing us to make use of the TT gauge, in which it was easy to show that there are two independent polarization modes of gravitational radiation. One can hence speculate that the same statement should be true when solving the linearized Einstein equation in the presence of sources. In this section, we want to show precisely that: We will find that it is always possible to decompose a small metric fluctuation into scalar, vector and tensor (SVT) dofs, out of which only two tensor dofs propagate.

Are there two propagating dofs only in vacuum?

To start with, let us consider the same decomposition $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ as before with $h_{\mu\nu}$ being sufficiently small in order to perform perturbative calculations. We can decompose $h_{\mu\nu}$ into parts that behave in the same way under rotations in three-dimensional Euclidean space, i.e. scalars, vectors and tensors: The component h_{00} transforms as a scalar, h_{0i} transforms as a vector and h_{ij} transforms as a spatial tensor. We can go even further and can decompose the vector field h_{0i} into a transverse and a longitudinal part. Similarly, the spatial tensor h_{ij} can further be decomposed into a diagonal part, a traceless part, an irrotational part and a remaining TT part:

$$h_{00} = 2\psi, \quad h_{0i} = \beta_i + \partial_i \gamma, \quad h_{ij} = -2\phi \delta_{ij} + \hat{P}_{ij} \lambda + \partial_{(i} \epsilon_{j)} + h_{ij}^{\text{TT}}. \quad (2.45)$$

For brevity of the notation, we introduced the projector $\hat{P}_{ij} = \partial_i \partial_j - \frac{1}{3} \delta_{ij} \Delta$, being traceless $\delta^{ij} \hat{P}_{ij} = 0$ by definition, and the symmetrization procedure $\partial_{(i} \epsilon_{j)} = \frac{1}{2} (\partial_i \epsilon_j + \partial_j \epsilon_i)$, which makes sure that the term is irrotational. The transversality of β_i and ϵ_i can be ensured through the two conditions $\partial_i \beta^i = \partial_i \epsilon^i = 0$. The transversality of $h_{\mu\nu}^{\text{TT}}$ corresponds to three conditions $\partial^i h_{ij} = 0$; the tracelessness gives yet another constraint $\delta^{ij} h_{ij}^{\text{TT}} = 0$, totaling in six constraints on the 16 components of $h_{\mu\nu}$. The 10 components of the symmetric $h_{\mu\nu}$ tensor field are now captured in the $10 = 4 \cdot 1 + 2 \cdot 2 + 1 \cdot 2$ components of the four scalar fields ψ, γ, ϕ and λ , the two transverse vector fields β_i and ϵ_i and the TT tensor field h_{ij}^{TT} .

Analogously, the energy-momentum tensor $T_{\mu\nu}$ can be decomposed like

$$T_{00} = \rho, \quad T_{0i} = S_i + \partial_i S, \quad T_{ij} = p \delta_{ij} + \hat{P}_{ij} \sigma + \partial_{(i} \sigma_{j)} + \sigma_{ij}^{\text{TT}}, \quad (2.46)$$

where ρ, S, p and σ are scalar fields, S_i and σ_i are transverse vector fields and σ_{ij}^{TT} is a TT tensor field. Also in this case, the six conditions $\partial^i S_i = \partial^i \sigma_i = \partial^i \sigma_{ij}^{\text{TT}} = \delta^{ij} \sigma_{ij}^{\text{TT}} = 0$ reduce the 16 dofs of $T_{\mu\nu}$ to 10, as required by $T_{\mu\nu}$ being symmetric.

These 10 dofs of the energy-momentum tensor are not independent. Energy-momentum conservation $\partial^\mu T_{\mu\nu} = 0$ yields 4 conditions, because of which only $6 = 10 - 4$ components out of $\rho, S, p, \sigma, S_i, \sigma_i$ and σ_{ij}^{TT} are independent. We will choose ρ, σ, S_i and σ_{ij}^{TT} to be the six independent components of $T_{\mu\nu}$ in the following. The same statement can be made about the fields $\psi, \gamma, \phi, \lambda, \beta_i, \epsilon_i$ and h_{ij}^{TT} on the left side of the Einstein equation based on the linearized Bianchi identity $\partial^\mu G_{\mu\nu} = 0$. However, the Einstein tensor $G_{\mu\nu}$ is a rather complicated function of those fields. We therefore take another road in this case and use the feature of linearized GR being invariant under a change of coordinates $x^\mu \rightarrow x^\mu + \xi^\mu$. Under this change of coordinates, the metric perturbation transforms as shown in the previous section in eq. (2.35). We now decompose $\xi_\mu = (\xi_0, \xi_i)$ into a scalar $\xi_0 = A$ and a vector with a longitudinal and transversal part $\xi_i = B_i + \partial_i C$, ensured by $\partial^i B_i = 0$.

*SVT decomposition
of the metric
perturbation*

*SVT decomposition
of $T_{\mu\nu}$*

$\partial^\mu T_{\mu\nu} = \partial^\mu G_{\mu\nu} =$
*0 removes four
dofs*

Under this change of coordinates, the introduced fields describing the metric perturbations transform as

$$\begin{aligned}\psi &\rightarrow \psi - \dot{A}, \quad \phi \rightarrow \phi + \frac{1}{3}\Delta C, \quad \gamma \rightarrow \gamma - A - \dot{C}, \quad \lambda \rightarrow \lambda - 2C, \\ \beta_i &\rightarrow \beta_i - \dot{B}_i, \quad \epsilon_i \rightarrow \epsilon_i - 2B_i \quad \text{and} \quad h_{ij}^{\text{TT}} \rightarrow h_{ij}^{\text{TT}}.\end{aligned}\quad (2.47)$$

As seen before in section 2.4.1, the TT tensor $h_{\mu\nu}^{\text{TT}}$ is indeed gauge-invariant. Instead of choosing a specific gauge as we did before, we will now construct the gauge-invariant Bardeen⁹ potentials

$$\Phi = -\phi - \frac{1}{6}\Delta\lambda, \quad \Psi = -\psi + \dot{\gamma} - \frac{1}{2}\ddot{\lambda}, \quad \Xi_i = \beta_i - \frac{1}{2}\dot{\epsilon}_i \quad \text{and} \quad h_{ij}^{\text{TT}}. \quad (2.48)$$

The Bardeen vector Ξ_i is transverse by construction since both β_i and ϵ_i are transverse as well. We are hence left with $6 = 1 + 1 + 2 + 2$ dofs which are left unchanged by a small shift ξ_μ of the coordinates. The computation of the Einstein equations is lengthy, as usual. We will therefore skip the derivation and only quote a selection of them here in order to interpret them in the context of propagating and non-propagating degrees of freedom:

$$\Delta\Phi = -\frac{\rho}{2m_{\text{Pl}}^2}, \quad \Delta\Psi = \frac{\rho - 2\Delta\sigma}{2m_{\text{Pl}}^2}, \quad (2.49a)$$

$$\Delta\Xi_i = -\frac{2S_i}{m_{\text{Pl}}^2} \quad \text{and} \quad \square h_{ij}^{\text{TT}} = -\frac{2\sigma_{ij}^{\text{TT}}}{m_{\text{Pl}}^2}. \quad (2.49b)$$

Of course, this set of equations is not closed and therefore not sufficient to describe the entirety of perturbation theory. In particular, next to the missing Einstein equations, there are conservation equations following from $\partial^\mu T_{\mu\nu} = 0$ which close the above system. Note, however, that only the tensor perturbations follow a wave equation. The full set of equations can be found in ref. [37].

Interpretation of the Bardeen potentials

The Bardeen scalar Φ can be interpreted as a Newtonian gravitational potential, only differing from $-\Psi$ through the scalar part σ of the anisotropic stress. The scalar metric perturbations Φ and Ψ follow Poisson equations, as expected from Newtonian gravity. The Bardeen vector instead represent the response of the metric to vorticity. The equation of motion for Ξ_i resembles the spatial part of the equation of motion $\Delta A^\mu = j^\mu$ of the four-potential A^μ in classical electrodynamics (in the Lorentz gauge $\partial_\mu A^\mu = 0$, which corresponds to our transversality condition $\partial^i \Xi_i = 0$). This third equation presents an extension of Newtonian gravity which is referred to as gravito-electromagnetism. It predicts phenomena like the experimentally tested Lense-Thirring effect, in which a rotating massive body drags the space-time around it. The last equation above shows that it is the transverse-traceless part of the anisotropic

⁹ We will refer to the quantities in eq. (2.47) as Bardeen potentials, even though they were first introduced only in the more general scenario of an SVT decomposition in an expanding background, which we will come to in section 2.4.4.

stress which sources the two independent polarization modes of GWs h_{ij}^{TT} . In the absence of sources, the Bardeen scalars and vector must vanish, whereas the GW part of h_{ij} instead survives, as previously found in section 2.4.1. Still, also in the presence of sources, we find that there are in general two modes in h_{ij}^{TT} which both follow a wave equation. The hypothesized statement from the beginning of this section hence holds true.

We summarize this mathematically challenging chapter as follows: A metric perturbation $h_{\mu\nu}$ generally contains spurious gauge dofs, physical but non-radiative dofs, as well as physical and radiative dofs. Due to the presence of non-radiative dofs, it is not always possible to write a metric perturbation in TT gauge. Only in the vacuum case, the TT gauge can be used, as otherwise the components T_{00} and T_{0i} could not be eliminated. Yet, the two dofs in $h_{\mu\nu}^{\text{TT}}$ are the only physical dofs that represent gravitational radiation, independent of the choice of reference frame and whether in vacuum or not.

*Also in matter
there are two
gauge-independent
propagating modes!*

2.4.3 Gravitational waves in a curved background

In the previous section we have identified GWs with the two propagating, gauge-invariant dofs of a generic small metric perturbation around a globally flat Minkowski background. In this section we drop the latter assumption and tackle the definition of GWs as propagating metric perturbations over a curved background. In fact, this step is necessary in order to define an energy-momentum tensor $t_{\mu\nu}$ of the GW. It is clear from the effect of gravitational radiation on test masses (see section 2.4.1) that they must carry energy. Otherwise they would not be able to displace mirrors in interferometers. However, to arrive at an expression for $t_{\mu\nu}$, one has to identify how GWs themselves curve spacetime. This simple statement already shows us that we need to go beyond linearized gravity. Instead, we want to explore how the background reacts to a metric perturbation.

*Going beyond the
flat background for
finding $t_{\mu\nu}$*

The task of splitting the metric $g_{\mu\nu}(x) = \bar{g}_{\mu\nu}(x) + \delta g_{\mu\nu}(x)$ into a background part \bar{g} and a perturbation δg with $|\delta g_{\mu\nu}(x)| \ll |\bar{g}_{\mu\nu}(x)|$ is far from being a simple task for a generic metric. In the general case, it is indeed not possible to precisely distinguish a background from the fluctuations on top of it. This ambiguity can be understood more intuitively when comparing with a more familiar scenario: water. Waves in the sea can of course be split into a part which arises due to the superposition of many incoherent waves and a single wave propagating on top of these, e.g. towards a beach. A precise distinction of the two modes, one defining the sea level and the other one being what we refer to as a wave in everyday life, however requires a separation of scales of their respective wavelengths.

*„Fischer, Fischer,
wie tief ist das
Wasser?“*

Separation of scales

A sub-Hubble GW is well-defined

Averaging over times works just as well

The energy-momentum tensor $t_{\mu\nu}$ of a gravitational wave

Mathematically speaking, it is possible to perform a sort of renormalization procedure in order to split off a high-frequency part of the metric, corresponding to the wave part $\delta g_{\mu\nu}$, from the low-frequency part, corresponding to the background $\bar{g}_{\mu\nu}$, which can be interpreted as a quasi-stationary Newtonian potential. Using this method, one can then integrate out short-wavelength modes by spatially averaging over the Einstein equation in a volume defined by some intermediate length scale. If both length scales coincide, a GW part cannot be unambiguously identified within the total metric $g_{\mu\nu}$.

Luckily, in the case of GWs propagating in an FLRW background, this separation of scales indeed exists for GWs with wavelengths being much smaller (or larger) than the Hubble radius. Indeed, a causal production mechanism for gravitational radiation is only able to produce sub-Horizon modes with a wavelength of $\lambda = 1/f \leq \frac{a_0}{a_* H_*}$ today, where a_* and H_* denote the scale factor and Hubble parameter at the production of the wave. The universe is expanding, meaning that $a_* H_* \gg a_0 H_0$, which immediately leads to $\lambda \ll H_0^{-1} \equiv L_{\text{bg}}$. A GW produced through a causal process at a given time will thus always have a wavelength that is much smaller than the current Hubble radius, which sets a length scale L_{bg} on which the background metric varies. We can therefore already anticipate that gravitational radiation can indeed be defined in a cosmological context.

For GWs arriving in the gravitational potential of Earth the situation is a little more complicated. Earth-based interferometers have their peak sensitivities in the kHz band, corresponding to wavelengths on the scale of 100 km. On these length scales, the deviations in the gravitational field $\delta g_{00} \sim 10^{-9}$ of the Earth exceed the characteristic strain amplitudes $\delta g_{ij} \sim 10^{-21}$ measured by interferometers by orders of magnitude. Luckily, the Earth gravitational field is quasi-static ($f_{\text{bg}} < 0.1$ Hz) compared to the kHz GWs LIGO detected in 2014. Performing an averaging over an intermediate time scale hence allows to find a sensitivity window which is safe from time variations in the Newtonian potential.

We will skip a rigorous treatment of the Einstein equations for a GW propagating in a curved background, referring the interested reader to the excellent discussion of this complex topic in chapter 1.4 of Maggiore's book [37]. Here, we will summarize the argument presented there: The Ricci tensor $R_{\mu\nu} = \bar{R}_{\mu\nu} + R_{\mu\nu}^{(1)} + R_{\mu\nu}^{(2)} + \mathcal{O}(\delta g^3)$ can be ordered into parts according to the powers of $\delta g \equiv \mathcal{O}(|\delta g_{\mu\nu}|)$ appearing in it, for a given perturbation $\delta g_{\mu\nu}$ which was split off from the quasi-static background metric $\bar{g}_{\mu\nu}$. Eventually, one finds that the energy-momentum tensor $t_{\mu\nu}$ corresponding to the effect of the metric perturbation $\delta g_{\mu\nu}$ on the background metric $\bar{g}_{\mu\nu}$ is hidden in the part of the Ricci tensor which is quadratic in δg , i.e. $R_{\mu\nu}^{(2)}$, which vanishes in linearized GR, and reads

$$t_{\mu\nu} = \frac{m_{\text{Pl}}^2}{4} \langle \partial_\mu \delta g_{\alpha\beta} \partial_\nu \delta g^{\alpha\beta} \rangle. \quad (2.50)$$

The derivatives in the above equation refer to the coordinates defined by the background metric and $\langle \cdot \rangle$ has to be understood as an averaging over an intermediate length scale $\lambda \ll \ell \ll L_{\text{bg}}$ or intermediate time scale $1/f \ll \tau \ll 1/f_{\text{bg}}$. In particular in the TT gauge the Lorentz condition $\partial^\mu \delta g_{\mu\nu} = 0$ implies that $t_{\mu\nu}$ only depends on the two physical modes in $h_{ij}^{\text{TT}} = \delta g_{ij}$, such that we can immediately read off the energy density associated to a GW,

$$\rho_{\text{gw}} = t_{00} = \frac{m_{\text{Pl}}^2}{4} \langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \rangle = \frac{m_{\text{Pl}}^2}{2} \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle. \quad (2.51)$$

Another important result of solving the Einstein equations after splitting off a slowly varying from a quickly varying part is that GWs propagate along null geodesics of the background metric, which immediately leads to the result that also gravitational radiation is subject to optical phenomena like gravitational lensing, absorption and scattering, even though their effect usually is negligible due to a suppression by the weakness of the gravitational interaction.

2.4.4 Gravitational waves in an expanding background

We will now combine our previous discussion of the homogeneous universe in section 2.3 with the weak-field limit of GR in which we showed that it is possible to perform an SVT decomposition. To do so we start by specifying the line element

$$ds^2 = a^2(\tau)(\eta_{\mu\nu} + h_{\mu\nu}) dx^\mu dx^\nu, \quad (2.52)$$

which corresponds to the perturbed FLRW metric in conformal time. The metric $g_{\mu\nu}(x)$ hence splits into a background part $\bar{g}_{\mu\nu}(x) = a^2(\tau)\eta_{\mu\nu}$ and perturbations $\delta g_{\mu\nu}(x) = a^2(\eta)h_{\mu\nu}(x)$. The conformal time parameter τ is related to the time parameter in the unperturbed FLRW metric in eq. (2.2) through $d\tau = dt/a(t)$. We introduce conformal time here in order to show that the following SVT decomposition closely parallels the one performed for a flat background in the previous section 2.4.2. Further, it is common to introduce the conformal Hubble parameter $\mathcal{H} = a'/a = aH = \dot{a}$, where we used that for any function $f(t)$ one has $f' = af$ with the prime denoting a derivative with respect to τ .

Cosmological perturbation theory

In order to understand the interplay of perturbations $h_{\mu\nu}(x)$ of the metric and perturbations of the energy momentum tensor, let us use the same decomposition into functions with different SVT helicities as in eq. (2.45). As above for a flat background, we find that also the perturbed FLRW metric can be decomposed into parts that transform as scalars ($\psi, \phi, \gamma, \lambda$), transversal vectors (β_i, ϵ_i) and a TT tensor ($h_{\mu\nu}^{\text{TT}}$) under spatial rotations. Under a linearized gauge transformation $x^\mu \rightarrow x^\mu + \xi^\mu$ the metric perturbation $\delta g_{\mu\nu}$ transforms as

SVT decomposition on an FLRW background

$$a^2 h_{\mu\nu} \rightarrow a^2 h_{\mu\nu} - (\nabla_\nu \xi_\mu + \nabla_\mu \xi_\nu) \quad (2.53)$$

*Gauge dependence
of the helicity
eigenfunctions*

*The Bardeen
potentials*

*Perturbations
around a perfect
fluid*

*The equation of
motion for the
Bardeen potentials*

with ∇_μ being a derivative with respect to the background metric as required by eq. (2.34). We can write the performed small coordinate shift again as $\xi^\mu = (\xi^0, \xi^i) = (-A, B^i + \partial^i C)$ with $B_i = B^i$ and $\partial_i C = \partial^i C$, such that $\xi_\mu = g_{\mu\nu} \xi^\nu = a^2 (A, B_i + \partial_i C)$. Under such a coordinate change, the SVT functions change as

$$\begin{aligned}\psi &\rightarrow \psi - (A' + \mathcal{H}A), & \phi &\rightarrow \phi + \left(\frac{1}{3} \Delta C - \mathcal{H}A \right), \\ \gamma &\rightarrow \gamma - (A + C'), & \lambda &\rightarrow \lambda - 2C, & \beta_i &\rightarrow \beta_i - B'_i, \\ \epsilon_i &\rightarrow \epsilon_i - 2B_i & \text{and} & & h_{\mu\nu}^{\text{TT}} &\rightarrow h_{\mu\nu}^{\text{TT}}.\end{aligned}\quad (2.54)$$

In the flat limit $a = 1$, this set of transformations reproduces the ones performed in eq. (2.47). Just as before, we do not specify a gauge but rather construct four gauge-invariant Bardeen potentials in order to reduce the amount of independent dofs from 10 to 6, which can again be motivated by having four independent components of the Bianchi identity:

$$\begin{aligned}\Phi &= -\phi - \frac{\Delta\lambda}{6} + \mathcal{H} \left(\gamma - \frac{\lambda'}{2} \right), & \Psi &= -\psi + \frac{1}{a} \frac{d}{d\tau} \left[a \left(\gamma - \frac{\lambda'}{2} \right) \right], \\ \Xi_i &= \beta_i - \frac{1}{2} \epsilon'_i & \text{and} & & h_{\mu\nu}^{\text{TT}} &.\end{aligned}\quad (2.55)$$

Again, in the limit $a = 1$, these simplify to the gauge-invariant Bardeen potentials introduced before in eq. (2.48). As above, the requirement that ξ_i is transversal immediately follows from the transversality of β_i and ϵ_i . The $6 = 1 + 1 + 2 + 2$ independent components of the above fields can hence be identified with the six physical metric components which are dynamical.

The left-hand side of the Einstein equation can now be parameterized as a function of the scale factor $a(\tau)$ and the Bardeen scalars, vector and tensor. We also recycle the previously used parameterization of the energy-momentum tensor from eq. (2.46), however splitting the energy density and pressure into a background and perturbation component, $\rho = \bar{\rho} + \delta\rho$ and $p = \bar{p} + \delta p$. Since raising and lowering indices now involves factors of the scale factors, it is most convenient to specify the components of T_ν^μ instead of $T_{\mu\nu}$,

$$\begin{aligned}T_0^0 &= -(\bar{\rho} + \delta\rho), & T_0^i &= S^i + \partial^i S & \text{and} \\ T_j^i &= (\bar{p} + \delta p) \delta_j^i + \left(\partial^i \partial_j - \frac{1}{3} \delta_j^i \Delta \right) \sigma + \frac{1}{2} (\partial^i \sigma_j + \partial_j \sigma^i) + \sigma_{ij}^{\text{TT}},\end{aligned}\quad (2.56)$$

Using this parameterization is particularly convenient because one solution of the Einstein equations already becomes obvious: For vanishing matter perturbations, only $\bar{\rho}$ and \bar{p} remain and the Friedmann eqs. (2.4) can be obtained from the diagonal elements (00) and (ii) of the Einstein equations (in conformal time and with the replacement $\rho \rightarrow \bar{\rho}$, $p \rightarrow \bar{p}$). Expressing the other Einstein equations in terms of the linearized metric

perturbations and the above matter perturbations requires a lengthy computation. For brevity, we only cite the emerging equations of motion for the Bardeen potentials here and refer to ref. [37] for the full set of equations:

$$\begin{aligned} \Delta\Phi - 3\mathcal{H}(\Phi' - \mathcal{H}\Psi) &= -2\frac{a^2\delta\rho}{m_{\text{Pl}}^2}, \quad \Delta(\Phi + \Psi) = -\frac{a^2\Delta\sigma}{m_{\text{Pl}}^2}, \\ \Delta\Xi_i &= -\frac{2a^2S_i}{m_{\text{Pl}}^2} \quad \text{and} \quad (h_{ij}^{\text{TT}})'' + 2\mathcal{H}(h_{ij}^{\text{TT}})' - \Delta h_{ij}^{\text{TT}} = \frac{2a^2\sigma_{ij}^{\text{TT}}}{m_{\text{Pl}}^2}. \end{aligned} \quad (2.57)$$

We can again see that these equations of motion closely resemble their counterparts in eq. (2.49b) for a flat background. In an expanding background, however, the Poisson equation for $\Delta\Phi$ becomes time-dependent. In fact, it allows for growing solutions, in which an initial overdensity keeps growing. This is precisely the origin of the growth of structure in the early universe and ultimately the reason we can see stars on the night sky. The vector equation remains unchanged (up to an overall factor a^2). The complete set of SVT-decomposed Einstein equations also includes the equation $\Xi'_i + 2\mathcal{H}\Xi_i = 2a^2\sigma_i/m_{\text{Pl}}^2$, which indicates the decay of vector modes in the absence of a vector part of anisotropic stress [85]. As vector modes further only satisfy a Poisson equation but not a wave equation, they do not propagate and thus only play a negligible role in cosmology. The equation for the tensor perturbation remains a wave equation also in the case of an expanding background metric. However, there now appears a damping term due to the expanding FLRW background. In the following section we want to see how this term referred to as Hubble friction leads to a damping of the GW amplitude over time.

*Structure grows,
vectors decay and
GWs propagate*

2.4.5 Equation of motion for a gravitational wave

The last equation in (2.57) is the equation of motion of a GW in the expanding FLRW background and thus central for this thesis. In order to see what effect the Hubble expansion has on an individual wave, it is convenient to rescale the metric perturbation to $H_{ij}(\mathbf{x}, \tau) = a(\tau)h_{ij}^{\text{TT}}(\mathbf{x}, t)$. Upon going to Fourier space we then obtain

$$\tilde{H}_{ij}'' + \left(k^2 - \frac{a''}{a}\right)\tilde{H}_{ij} = \frac{2a^3}{m_{\text{Pl}}^2}\tilde{\sigma}_{ij}^{\text{TT}}. \quad (2.58)$$

In conformal time, the scale factor changes as $a(\tau) \propto \tau^n$ with $n = 1$ in radiation domination, $n = 2$ in matter domination and $n = -1$ during inflation. In particular, this means that after its emission a GW follows the equation $\tilde{H}_{ij}'' + k^2\tilde{H}_{ij} = 0$ in radiation domination.¹⁰ This equation of motion is easily solved by $\tilde{H}_{ij} = A_{ij}(\mathbf{k})\sin(k\tau) + B_{ij}(\mathbf{k})\cos(k\tau)$.

*A GW is also just
a harmonic
oscillator*

¹⁰ It is not precisely true that after its emission a GW evolves as in vacuum. In fact, there is an effect of relativistic species moving along the same geodesics of the background metric as the GW [37, 86]. This generates an anisotropic stress $\sigma = \mathcal{O}(h)$

Using that for sub-Hubble modes $H'_{ij} \sim k H_{ij} \gg \mathcal{H} H_{ij}$, we can express the term contributing to the energy density in eq. (2.51) as

$$\dot{h}_{ij}^{\text{TT}} \dot{h}_{\text{TT}}^{ij} = \frac{1}{a^4} (H'_{ij} H^{ij\prime} + 2\mathcal{H} H'_{ij} H^{ij} + \mathcal{H}^2 H_{ij} H^{ij}) \simeq \frac{H'_{ij} H^{ij\prime}}{a^4}. \quad (2.59)$$

The energy density of the GW under consideration hence reads

$$\rho_{\text{gw}}(\tau) = \frac{m_{\text{Pl}}^2}{4a^4} \langle H'_{ij} H^{ij\prime} \rangle = \frac{m_{\text{Pl}}^2}{4a^4} \int \frac{d^3k}{(2\pi)^3} \tilde{H}'_{ij}(\mathbf{k}, \tau) \tilde{H}^{ij\prime*}(\mathbf{k}, \tau). \quad (2.60)$$

In the last step we expanded each $H_{ij}(\mathbf{x}, \tau)$ into Fourier components $\tilde{H}_{ij}(\mathbf{k}, \tau)$, used that the spatial average can be performed through evaluation of the integral $\int d^3x \exp[-i(\mathbf{k} + \mathbf{k}') \cdot \mathbf{x}] = (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{k}')$, which in turn eliminates the integral over d^3k' , and set $\tilde{H}'_{ij}(-\mathbf{k}, \tau) = \tilde{H}'_{ij}^*(\mathbf{k}, \tau)$ in order for $h_{ij}(\mathbf{x}, \tau)$ to be real. Note that the remaining integrand is a function oscillating over a short timescale $1/k$. As we are only interested in the evolution of ρ_{gw} over cosmic timescales, we average over these oscillations. Using that $\langle \sin^2 x \rangle = \langle \cos^2 x \rangle = \frac{1}{2}$ and $\langle \sin x \cos x \rangle = 0$ we eventually obtain

$$\rho_{\text{gw}}(\tau) = \frac{m_{\text{Pl}}^2}{8a^4(\tau)} \int \frac{d^3k}{(2\pi)^3} k^2 [A_{ij}^*(\mathbf{k}) A^{ij}(\mathbf{k}) + B_{ij}^*(\mathbf{k}) B^{ij}(\mathbf{k})]. \quad (2.61)$$

Gravitational radiation scales like radiation

We hence find that once a GW is emitted, its energy density indeed follows $\rho_{\text{gw}} \propto a^{-4}$ as expected for radiation. The integral of the right-hand side is independent of time but remains not being very insightful. In section 2.4.7 we will consider the case of a stochastic GWB instead of a single plane GW as in the above calculation. Expanding over plane waves, we will find that the integrand on the right hand side can be interpreted as the power spectrum of GWs.

Note that in our simplified calculation above, we used that the term a''/a vanishes in radiation domination. However, as long as $a''/a \ll k$ our above calculation remains valid also for other equation of state parameters $\bar{p}/\bar{\rho}$. This is in particular the case for sub-Hubble modes in any generic epoch [37]. The main effect of a varying equation of state parameter are spectral distortions of modes which enter the Hubble sphere only late [86, 87]. As we only consider modes that are already deep inside the horizon when the equation of state parameter drops during radiation-matter equality at $z \sim 3600$, these spectral distortions are of no importance for this thesis. This in turn implies that we do not need to solve the equations of motion of a metric perturbation, but can rely on the simple $\rho_{\text{gw}} \propto a^{-4}$ scaling in the following calculations.

We do not need to solve the equations of motion each time

also after the initial source of the GW has become inactive. In particular, free-streaming neutrinos damp modes that enter the Hubble sphere between neutrino decoupling and matter-radiation equality.

2.4.6 Stochastic GW signals from the early cosmos

GWs from the early universe are typically sourced by many individual, incoherently acting contributions to the anisotropic stress. Examples of these include bubble collisions in cosmological PTs (see chapter 4), oscillating cosmic string loops and localized features called kinks and cusps propagating along these loops, as well as decaying domain walls. Further, going to second order in cosmic perturbation theory, one finds that the decomposition into decoupled SVT components of the metric breaks down. Scalar perturbations Φ and Ψ can then be shown to also contribute to the source term σ_{ij}^{TT} , e.g. through a term proportional to $\partial_i \Phi \partial_j \Psi$ [11]. Also cosmic inflation is predicted to give rise to an irreducible background of GWs through the amplification of quantum fluctuations and possibly due to reheating.

Early-universe sources of GWs

Either of these sources active in the early cosmos can emit a potentially observable signal. However, unlike the GWs observed as transient signals by the LVK collaboration since 2014, instead a stochastic GWB would be detected. This means that it is not possible to precisely predict $h_{ij}^{\text{TT}}(\mathbf{x}, t)$ given an early universe phenomenon emitting GWs. Instead, h_{ij}^{TT} is a random variable whose statistical distribution can be predicted. In particular, the variance of this distribution is relevant, as it is related to the energy density carried by the GW, as shown in section 2.4.3. Obviously, there is only one realization of each event that once happened in the early universe. By using a variant of the ergodic hypothesis, the ensemble average can however be interpreted as an averaging procedure over large enough portions of space and time.

In practice, in order to predict a GW spectrum, one typically runs numerical simulations of a given physical process in the early universe to obtain σ_{ij}^{TT} . One then integrates the equations of motions (2.58) up to the point in time when the source ceases to be active and identifies the GW spectrum at the moment it decouples and starts propagating it as a free wave. The necessary ensemble averaging can be performed by integrating over modes in a large enough simulated volume, as demonstrated in section 2.4.5. Since the metric perturbation at a given coordinate point is random, also in order to observe a stochastic GWB as a form of irreducible noise, one eventually needs to average over a large enough portion of space for a long enough time.

In order to make use of the ergodic hypothesis, two requirements have to be met, however: First, the process sourcing the GW production needs to have the same initial conditions at every point in space. This condition is usually fulfilled due to the homogeneity and isotropy of the FLRW background in the case of cosmological sources. The second condition is causality: We further need to require that the process sourcing the

The ergodic hypothesis

GW is causal (i.e., it acts within a causal horizon) and that the Hubble sphere was smaller than today's Hubble sphere [11].¹¹

We can combine both requirements in order to obtain an estimate of the correlation scale of a GW signal from the primordial universe. For a process happening at the same instance throughout the universe and it being causal, its largest possible correlation scale ℓ_* at the time of production is roughly a Hubble radius H_*^{-1} , where an asterisk refers to the time of production. The corresponding correlation length scale redshifted to today reads $\ell_0 = \ell_* \frac{a_0}{a_*} \lesssim H_*^{-1} \frac{a_0}{a_*}$. In order to calculate the ratio of scale factors, we can employ the conservation of entropy in the SM bath requiring that $h_{\text{SM}}(T) T^3 a^3(T)$ is constant in time (assuming that there is no intermediate entropy injection from a DS). Plugging in $T_0 = 2.4 \cdot 10^{-13} \text{ GeV}$ and $h_{\text{SM}}^0 = 3.9$, we obtain

$$\frac{a_0}{a_*} = \left(\frac{h_{\text{SM}}^*}{h_{\text{SM}}^0} \right)^{1/3} \left(\frac{T_*}{T_0} \right) \simeq 1.3 \cdot 10^{13} \left(\frac{h_{\text{SM}}^*}{100} \right)^{1/3} \left(\frac{T_*}{\text{GeV}} \right). \quad (2.62)$$

We want to compare the correlation length ℓ_0 redshifted to today with today's Hubble radius H_0^{-1} . The latter is related to the Hubble radius H_*^{-1} at emission of the GW signal through

$$H_* = H_0 \sqrt{\sum_i \Omega_i \frac{\rho_i(T_*)}{\rho_i^0}} \simeq H_0 \sqrt{\Omega_{\text{rad}} \frac{\rho_{\text{rad}}(T_*)}{\rho_{\text{rad}}^0}}. \quad (2.63)$$

In the first step we employed the Friedmann eq. (2.4), whereas in the second step it was assumed that the emission happens in the early universe in radiation domination, i.e. $T_* \gtrsim 1 \text{ eV}$. The ratio of radiation energy densities reads $\rho_{\text{rad}}(T_*)/\rho_{\text{rad}}^0 = g_{\text{SM}}^* T_*^4 / (g_{\text{SM}}^0 T_0^4)$. Using entropy conservation another time in order to express the ratio of temperatures as a ratio of scale factors then immediately yields the wanted relation

$$\ell_0 H_0 \lesssim 1.3 \cdot 10^{-11} \left(\frac{100}{g_{\text{SM}}^*} \right)^{1/6} \left(\frac{\text{GeV}}{T_*} \right), \quad (2.64)$$

where in the last step we used that $g_{\text{SM}}^* \simeq h_{\text{SM}}^*$ for sufficiently high temperatures. This result clearly shows that any GW signal produced by a causal process in a short period of time around temperature T_* will have tiny correlation lengths today compared to the Hubble radius. Analogously to the above computation one can also compute the number of regions in the sky in which a primordial GW would hypothetically be sourced by a single realization of an early universe process [11]. The GWB from a model turning the EWPT first-order would for instance be sourced by a superposition of sources from *at least* 10^{24} uncorrelated regions of the celestial sphere. Hence, a cosmological GW signal can in practice only be observed through the determination of the statistical distribution of tensor fluctuations h_{ij}^{TT} . Pictorially speaking, a cosmic

Cosmic GWs can only be measured as noise

¹¹ The latter condition is not fulfilled in the case of inflationary GWs. Due to the stochastic nature of the quantum fluctuations of the inflaton field, which then become macroscopic during inflation, the emitted GWB is still stochastic.

GWB will hence manifest itself as a sort of noise an observatory cannot get rid of, much like Penzias and Wilson were not able to eliminate the CMB noise from their Horn antenna [88].

2.4.7 The spectrum of primordial gravitational waves

The central result of the last section 2.4.6 is that a GW signal from the early cosmos will necessarily be a superposition of many individual sources' emitted radiation. We now want to connect this with our previous discussion on the energy density of a single plane wave in section 2.4.5 in order to define the spectrum of a cosmic GWB. Employing the TT gauge we can expand the solution of the equation of motion (2.58) as a superposition of plane waves moving in direction $\hat{\mathbf{n}} = \mathbf{k}/|\mathbf{k}|$

$$h_{ij}(t, \mathbf{x}) = \sum_{A=+, \times} \int_{-\infty}^{+\infty} df \int d^2\hat{\mathbf{n}} \tilde{h}_A(f, \hat{\mathbf{n}}) e_{ij}^A(\hat{\mathbf{n}}) e^{-2\pi if(t - \hat{\mathbf{n}} \cdot \mathbf{x})}. \quad (2.65)$$

Superposition of plane waves

The polarization vectors therein can be expressed as $e_{ij}^+(\hat{\mathbf{n}}) = \hat{\mathbf{u}}_i \hat{\mathbf{u}}_j - \hat{\mathbf{v}}_i \hat{\mathbf{v}}_j$ and $e_{ij}^\times(\hat{\mathbf{n}}) = \hat{\mathbf{u}}_i \hat{\mathbf{v}}_j - \hat{\mathbf{v}}_i \hat{\mathbf{u}}_j$, where $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are unit vectors orthogonal to the propagation direction of the GW and each other. The Fourier amplitudes $\tilde{h}_A(f, \hat{\mathbf{n}}) = \tilde{h}_A^*(-f, \hat{\mathbf{n}})$ follow a statistical distribution. We will now show how for a stationary, Gaussian, isotropic and unpolarized background the whole distribution is determined through the variance of $\tilde{h}(f)$, which we then identify with the energy spectrum of the GWB.

STATIONARITY implies that the two-point correlator $\langle h_A(t) h_{A'}(t') \rangle$ only depends on the time difference $t - t'$ and not separately on the two cosmic times t and t' . Since observations of GWs take place on time scales $t - t' < \mathcal{O}(10 \text{ yr})$ much shorter than the age of the signals $t, t' = \mathcal{O}(14 \text{ Gyr})$, this assumption is usually very well justified. An exception are sources which keep being active over long enough time scales, see ref. [89]. Stationarity also leads to $\langle h_A(t) \rangle$ being approximately constant on human time scales, thereby contributing marginally to the relative density of vacuum energy ρ_Λ . This is because $\langle h_A \rangle = \text{const}$ leads to a stationary curvature of spacetime, analogously to the case of a cosmological constant. We are only interested in the time-dependent part of metric fluctuations and can thus ignore the first moment of the distribution by setting $\langle h_A \rangle = 0$ in our analysis.

Old GWBs are stationary

GAUSSIANITY suggests that the full statistical information one can possibly obtain from the background is contained in its mean, which we set to zero, and its variance $\langle h_A(t) h_{A'}(t') \rangle$. All higher moments can then be computed from the two-point function. This assumption is supported by our argument from section 2.4.6 that a single signal today is sourced by a large number of independent processes. The central limit theorem

The central limit theorem also holds for GWBs

then requires cosmic backgrounds to be Gaussian. Also the irreducible GWB from quantum fluctuations during inflation is Gaussian, in that case, however, rather due to the inflaton field being a free quantum field whose amplitudes follow a Gaussian probability distribution. A small slow-roll suppressed amount of non-Gaussianity can, however, be present in that case due to deviations from a perfect de Sitter-like inflationary phase. Note that the assumption of Gaussianity also needs to be dropped for astrophysical GWBs being formed by a superposition of only a few sources. In that case, further information on the distribution of the metric perturbations can be obtained from higher-point correlators.

Anisotropy hints towards an astrophysical GWB origin

ISOTROPY is a very well justified assumption for cosmic GWBs for processes that happen at the same point in time in different Hubble patches. Analogously to the CMB temperature anisotropies, which are of the order $\Delta T/T \sim 10^{-5}$, also a GWB should show no evident (anti-)correlations between the signal amplitude in different directions. We hence expect $\langle \tilde{h}_A^*(f, \hat{\mathbf{n}}) \tilde{h}_{A'}(f', \hat{\mathbf{n}}') \rangle$ to be proportional to $\delta(\phi - \phi') \delta(\cos \theta - \cos \theta')$, where ϕ and θ are the two angles in polar coordinates which specify the direction $\hat{\mathbf{n}}$, and where the constant of proportionality cannot depend on the direction $\hat{\mathbf{n}}$. Astrophysical backgrounds are typically much more anisotropic than cosmic backgrounds due to their strong correlation with the matter distribution in our galaxy compared to the isotropy of the primordial plasma. The detection of anisotropy therefore is a strong hint towards a local, astrophysical origin of a given signal.

NO POLARIZATION in the GWB is expected regardless if the signal is of astrophysical or cosmological origin, again due to the large number of individual sources. This means, that $\langle \tilde{h}_A^*(f, \hat{\mathbf{n}}) \tilde{h}_{A'}(f', \hat{\mathbf{n}}') \rangle$ is proportional to $\delta_{AA'}$, where the constant of proportionality is independent of the polarization index A .

Combining these four well-justified assumptions in the case of cosmological GWBs, we find that the two-point correlation function in Fourier space can be expressed as

$$\begin{aligned} \langle \tilde{h}_A^*(f, \hat{\mathbf{n}}) \tilde{h}_{A'}(f', \hat{\mathbf{n}}') \rangle &= \frac{1}{4\pi} \delta_{AA'} \delta(f - f') \delta(\phi - \phi') \\ &\quad \times \delta(\cos \theta - \cos \theta') \frac{1}{2} S_h(f), \end{aligned} \tag{2.66}$$

The spectrum is fully described by $S_h(f)$

where the full frequency dependence was absorbed into the polarization- and direction-independent two-sided spectral density $S_h(f) = S_h^*(-f)$. The factor $(4\pi)^{-1}$ cancels the normalization of the $d^2\hat{\mathbf{n}} = d\cos \theta d\phi$ integration; the factor $\frac{1}{2}$ instead was chosen in order to be consistent with the definition of one-sided noise spectral densities. Going back from Fourier space to position space and using that the polarization

tensors satisfy $e_{ij}^A e_{A'}^{ij} = 2\delta_{AA'}$, such that $\sum_A e_{ij}^A e_{A'}^{ij} = 4$, we find that the two-point correlator of $h_{ij}^0 \equiv h_{ij}(t_0, \mathbf{x} = 0)$ reads

$$\langle h_{ij}^0 h_0^{ij} \rangle = 4 \int_0^\infty df S_h(f). \quad (2.67)$$

Note that $S_h(f)$ only describes the current spectrum of the GW spectrum. We will discuss in the following section, how the spectrum redshifts between production and detection. The spectral density $S_h(f)$ has units of Hz^{-1} and is often used when comparing with experimental noise spectra in interferometers. Another way of specifying the spectrum is through

$$h_c^2(f) = 2f S_h(f) \quad \text{such that} \quad \langle h_{ij}^0 h_0^{ij} \rangle = 2 \int d\log f h_c^2(f), \quad (2.68)$$

which is useful as it is dimensionless and directly related to the order of magnitude of the tensorial metric perturbation, giving h_c the name characteristic strain amplitude.

In cosmology, the most frequently used way of describing a GW spectrum is instead based on the corresponding relic energy density of the GWB. One introduces the energy density in GWs normalized to the critical energy density $\rho_c^0 = 3m_{\text{Pl}}^2 H_0^2$

$$\Omega_{\text{gw}} = \int d\log f \frac{1}{\rho_c^0} \frac{d\rho_{\text{gw}}^0(f)}{d\log f} = \int d\log f \Omega_{\text{gw}}(f) \quad (2.69)$$

for which one needs to accept a slight abuse of notation in which $\Omega_{\text{gw}} \neq \Omega_{\text{gw}}(f)$ and $\rho_{\text{gw}}^0 \neq \rho_{\text{gw}}^0(f)$. Again, one obtains a dimensionless spectrum $\Omega_{\text{gw}}(f)$, now however with a related relative energy density Ω_{gw} , which can be compared to the relative energy density in radiation Ω_{rad} for instance. In particular for a spectrum $\Omega_{\text{gw}}(f)$ peaked around f_p , stemming for instance from a cosmic PT, one obtains that $\Omega_{\text{gw}}(f_p) \simeq \Omega_{\text{gw}}$, further motivating the use of this quantity in cosmology. In order to obtain the energy density of the plane wave-expanded spectrum and to relate it to $h_c(f)$ and $S_h(f)$, we plug eq. (2.65) into eq. (2.51), equate the ensemble average with a spatial average under use of the ergodic hypothesis as in section 2.4.5, and eventually obtain

$$\rho_{\text{gw}} = m_{\text{Pl}}^2 \int d\log f f (2\pi f)^2 S_h(f). \quad (2.70)$$

The factor $(2\pi f)^2$ therein comes from the two time derivatives in eq. (2.51) acting on the exponential of the Fourier transformation. The spectral energy density hence relates to the previous two spectral functions through

$$h^2 \Omega_{\text{gw}}(f) = \frac{4\pi^2}{3H_{100}^2} f^3 S_h(f) = \frac{2\pi^2}{3H_{100}^2} f^2 h_c^2(f), \quad (2.71)$$

The characteristic strain $h_c(f)$

The spectrum $\Omega_{\text{gw}}(f)$

Translating between spectra

where we introduce the quantity $H_{100} = 100 \text{ km/s/Mpc}$ and multiplied with $h \equiv H_0/H_{100} \simeq 0.69$ in order to make the conversion independent of the precise value of the Hubble constant H_0 , still being subject of the previously mentioned Hubble tension. Numerically, one can hence translate from $h^2\Omega_{\text{gw}}(f)$ to the other spectra using

$$S_h(f) \simeq 10^{-36} \left(\frac{\text{Hz}}{f} \right)^3 \frac{h^2\Omega_{\text{gw}}(f)}{\text{Hz}} \quad \text{and} \quad (2.72)$$

$$h_c(f) \simeq 10^{-18} \left(\frac{\text{Hz}}{f} \right) \sqrt{h^2\Omega_{\text{gw}}(f)}. \quad (2.73)$$

2.4.8 The redshift of gravitational wave spectra

We already discussed in the introduction of this thesis, that GWs decouple as early as a Planck time from the rest of the universe. The weakness of the gravitational interaction further allows us to describe gravitational radiation, once emitted, to a very good approximation through linearized gravity. The interaction of GWs with the remaining plasma and with itself can hence be neglected such that we can treat it as a freely propagating wave in an expanding background, as described by the equation of motion (2.58). We found that the corresponding energy density redshifts like $\rho_{\text{gw}} \propto a^{-4}$ as expected for radiation. Further, an individual mode scales with $f \propto a$ and the critical energy density is proportional to $\rho_c \propto H^2$. The relation between the spectrum at emission $\Omega_{\text{gw}}^*(f)$ and today $\Omega_{\text{gw}}(f)$ is therefore given by

$$\Omega_{\text{gw}}(f) = \mathcal{R}\Omega_{\text{gw}}^* \left(\frac{a_0}{a_*} f \right) \quad \text{with} \quad \mathcal{R} \equiv \left(\frac{a_*}{a_0} \right)^4 \left(\frac{H_*}{H_0} \right)^2. \quad (2.74)$$

The spectrum redshifts as expected for radiation...

...as long as the modes are sub-Hubble

Note that the above equation only holds for modes that are already sub-Hubble at the time of their production ($k \gg a_* H_*$). For lower-frequency modes of the GW spectrum, one generally expects a spectral distortion from entering at a time when the equation of state parameter has changed from the one at the time of production. This effect will however not be of relevance in the present thesis.

For getting a feeling of the orders of magnitude of the different quantities involved in eq. (2.74), we can use the expression for the scale factor ratio from eq. (2.62) and the ratio of Hubble parameters from eq. (2.63) to obtain

$$\mathcal{R}h^2 \simeq 1.6 \cdot 10^{-5} \left(\frac{100}{g_*(T_*)} \right)^{1/3}. \quad (2.75)$$

Each GW frequency corresponds to a point in time after the Big Bang

A given mode with physical wave number $a_0 k / a_*$ at production can be related to the Hubble rate H_* at that time to obtain $x_k = a_0 k / (a_* H_*)$. The corresponding frequency of that mode today is hence given by

$$f = \frac{1}{2\pi} \frac{k}{a_0} = \frac{x_k}{2\pi} \frac{a_*}{a_0} H_* \simeq 26 \text{ nHz} x_k \left(\frac{g_*(T_*)}{100} \right)^{1/6} \left(\frac{T_*}{\text{GeV}} \right). \quad (2.76)$$

Since causality requires $x_k \geq 1$, the above equation yields a lower bound on the frequency of a GWB caused by a process operative at temperature T_* . For the same reason, one can also predict that for processes during which the sources are active only over a brief period of time, the GW spectrum will be peaked. Vice, versa, in the opposite case of a long-lasting source, one rather expects an extended plateau in $\Omega_{\text{gw}}(f)$. For instance, the spectrum emitted by bubble collisions in a first-order phase transition (FOPT) typically peaks at a frequency given by eq. (2.76) with $x_k \simeq 10 - 10^4$, depending on the bubble size. Oscillating cosmic strings instead have a GW spectrum with a broad plateau spanning many orders of magnitude in frequency due to the emission happening over long time scales. The irreducible background in tensor fluctuations from inflation is also almost flat, in this case because the source is active while $a_* H_*$ grows exponentially.

Peaked and plateau-shaped spectra

Note that entropy conservation in the SM bath was assumed in deriving eqs. (2.62) and (2.63), which we reused here. In the presence of an entropy injection into the SM bath, for instance from a DS decay, this assumption breaks down and requires a correction of the above expression for $\mathcal{R}h^2$. We will discuss this correction in chapter 4.3.2.

A DS decay can change the redshift relation

2.4.9 Primordial GWB searches

We want to conclude this chapter with a brief discussion of the existing and predicted constraints on the cosmic GWB, the evidence for a background at nHz frequencies and the ΔN_{eff} bound on cosmological backgrounds. Fig. 2.8 shows an overview of these different sources of information. For the sake of simplicity we decided to only plot the projected sensitivity curves of LISA, the Big Bang Observatory (BBO) and the Einstein Telescope (ET), as well as the six lowest Fourier frequencies of the free-spectral analysis of the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) 15yr data set. For a more complete overview of the existent and prospective constraints see refs. [25, 90] for instance.

A selection of observatories

Given an effective noise spectrum $\Omega_{\text{eff}}(f)$ of an observatory, which is understood to include all relevant noise contributions from the detector but also from unresolved astrophysical backgrounds, we can construct the optimal-filter cross-correlated signal-to-noise ratio (SNR) for a cosmological GWB

$$\rho^2 = 2 t_{\text{obs}} \int_{f_{\min}}^{f_{\max}} df \left[\frac{\Omega_{\text{gw}}(f)}{\Omega_{\text{eff}}(f)} \right]^2. \quad (2.77)$$

The SNR of an interferometer

The factor t_{obs} denotes the duration of the observation and the observatory's frequency band spans f_{\min} to f_{\max} . The initial factor of 2 only appears when the SNR refers to a cross-correlation between different detectors. If instead an auto-correlation is used by the observatory, this factor 2 drops out of the equation. We will refer to a GW signal as

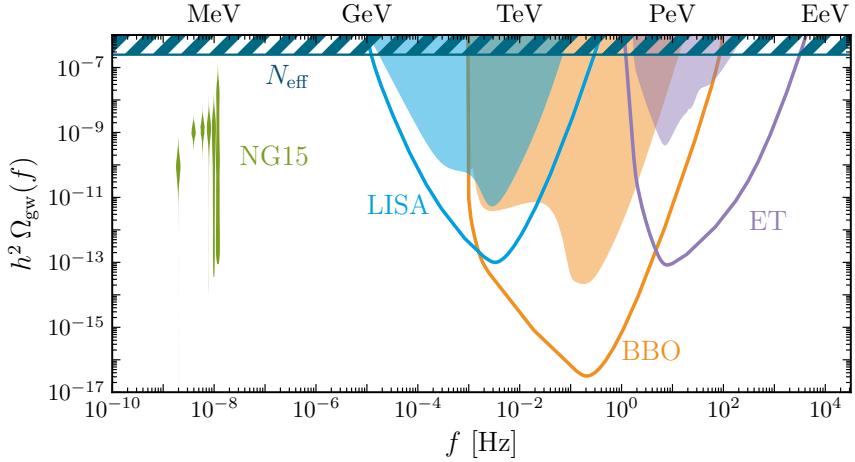


Figure 2.8: Overview of observational bounds on primordial GWBs. The blue, orange and purple filled areas correspond to the expected effective noise spectra $h^2 \Omega_{\text{eff}}$ of the interferometers LISA, BBO and ET from ref. [25]. The dips in the spectra correspond to the expected unresolved astrophysical background from galactic and extra-galactic white dwarf binaries. The blue, orange and purple contours are the associated PLI curves. The green violins depict the six lowest Fourier modes of the free-spectral analysis of the NANOGrav 15yr data set. The blue hatched band corresponds to the N_{eff} bound on primordial GWBs from eq. (2.82).

Experiment	Frequency range	ρ_{thr}	t_{obs}	Auto-correlated?
LISA	$10^{-5} - 1 \text{ Hz}$	10	4 yrs	✓
BBO	$10^{-3} - 10^2 \text{ Hz}$	10	4 yrs	✗
ET	$1 - 10^4 \text{ Hz}$	5	5 yrs	✓

Table 2.2: Threshold SNRs ρ_{thr} of future GW observatories taken from ref. [25].

observable if its SNR exceeds a given specific threshold, $\rho \geq \rho_{\text{thr}}$, which we list in table 2.2.

PLI curves

If $\Omega_{\text{gw}}(f)$ follows a power-law, it is convenient to not show $\Omega_{\text{eff}}(f)$ in plots as a comparison, but instead use so-called power-law integrated (PLI) curves $\Omega_{\text{PLI}}(f)$. For a given observatory, they are defined as

$$\Omega_{\text{PLI}}(f) = \max_b \left[\Omega_b^{\text{thr}} \left(\frac{f}{\bar{f}} \right)^b \right], \quad (2.78)$$

where \bar{f} is an arbitrary pivot frequency and

$$\Omega_b^{\text{thr}} \equiv \frac{\rho_{\text{thr}}}{\sqrt{2 t_{\text{obs}}}} \left[\int_{f_{\min}}^{f_{\max}} df \left(\frac{(f/\bar{f})^b}{\Omega_{\text{eff}}(f)} \right)^2 \right]^{-\frac{1}{2}} \quad (2.79)$$

is the minimum signal amplitude needed in order to reach the observatory's SNR threshold ρ_{thr} for a power-law shaped GWB $\Omega_{\text{gw}}(f) =$

$\Omega_b \left(\frac{f}{f}\right)^b$. If a given power-law shaped GWB spectrum intersects with the PLI spectrum, it can be considered observable by the respective observatory. However, it's important to note that this interpretation is only indicative if the signal does not actually follow a power-law shape. In such cases, comparing the SNR from eq. (2.77) with the threshold value provided in table 2.2 is necessary to reach a definitive conclusion.

Fig. 2.8 also shows the so-called Bayesian spectrograms (commonly referred to as violins) for the NANOGrav 15yr data set. We will discuss the precise origin and interpretation of these in detail in chapter 3. An easy way of interpreting them is as data points whose error bars have been upgraded to a full statistical distribution which is depicted as the violin's width. Note, however, that this interpretation can be misleading as it forgets about the Bayesian character of these distributions: The vanishing width of a given violin at low signal amplitudes does not indicate that it is impossible for the signal to have an amplitude lower than indicated by the violins. Instead, the range of the violins rather indicates the boundaries of their underlying prior ranges.

In fig. 2.8 we also show ΔN_{eff} bounds on primordial GWBs. In section 2.3.6 we introduced the quantity ΔN_{eff} which parameterizes any additional energy density ρ_{extra} in units of neutrino dofs after their decoupling. Both BBN and the CMB observables can be used to infer bounds on the Hubble rate around the temperature of neutrino decoupling and recombination, which can then be expressed in terms of bounds on ΔN_{eff} . If one equates the energy density ρ_{extra} in eq. (2.30) with the energy density ρ_{gw} of a primordial GWB, one can obtain a bound on its amplitude.

A note on violins

The N_{eff} bound on primordial GWBs

Using as reference values the 95% C.L. upper limit on N_{eff} from BBN and CMB combined from eq. (2.33) [74], $N_{\text{eff}} = 2.941 + 0.143 = 3.084$, and the corresponding $\Delta N_{\text{eff}} = 3.084 - 3.044 = 0.040$ as a normalization, we obtain

$$\left. \frac{\rho_{\text{gw}}}{\rho_\gamma} \right|_{t_{\text{BBN}}} = \left. \frac{\rho_{\text{gw}}}{\rho_\gamma} \right|_{t_0} < 9.1 \cdot 10^{-3} \left(\frac{N_{\text{eff}} - 3.044}{0.040} \right). \quad (2.80)$$

To arrive at this conclusion we used that after the e^+e^- -annihilation the photon energy density redshifts precisely like the GW energy density until today. Of course, this bound can only depend on the energy density of a primordial signal produced before $T \sim 0.1$ MeV, corresponding to the onset of BBN after the deuterium bottleneck. In particular, astrophysical backgrounds are therefore not affected by this bound. It should further be noted, that this bound can only depend on modes that already entered the Hubble sphere at that time. This corresponds to GWs which today have frequencies above $f_{\text{BBN}} \simeq 1.5 \cdot 10^{-11}$ Hz. To arrive at this value we used eq. (2.76) and for definiteness put in $x_k = 10$, requiring that only modes well within the Hubble sphere contribute. Using $h^2 \Omega_\gamma = 2.473 \cdot 10^{-5}$, we can equivalently write

$$\int_{f_{\text{BBN}}}^{\infty} \frac{df}{f} h^2 \Omega_{\text{gw}}(f) < 2.5 \times 10^{-7} \left(\frac{N_{\text{eff}} - 3.044}{0.040} \right). \quad (2.81)$$

N_{eff} puts bounds on the integrated spectrum

If we further assume that the signal is peaked over at least an order of magnitude in frequency (which is typically the case), we can further impose

$$h^2 \Omega_{\text{gw}}(f) \lesssim 2.5 \times 10^{-7} \left(\frac{N_{\text{eff}} - 3.044}{0.040} \right) \quad \text{for } f > f_{\text{BBN}}, \quad (2.82)$$

which is the BBN bound shown in fig. 2.8.

No sensitivity to primordial GWs at high frequencies

There are considerable ongoing efforts to extend the probed GW spectrum to higher frequencies [91, 92]. These proposals for instance recommend the use of the inverse Gertsenshtein effect, i.e. the coupling of GWs to electromagnetic fields, or the direct mechanical coupling of radio frequency cavities to GWs for their detection. So far, the existing limits are far from exceeding the ΔN_{eff} bound. We therefore do not show any further bounds at higher frequencies.

The upper horizontal axis in fig. 2.8 corresponds to the temperature T_* a given frequency refers to according to eq. (2.76), where we set $x_k = 1$ for definiteness and ignored the mild dependence of $g_* \simeq \mathcal{O}(100)$ on temperature. It hence becomes apparent that the PTA signal hints towards strong dynamics in the primordial plasma at MeV temperatures, whereas GW observatories sensitive to higher frequencies will be able to probe the primordial plasma at even higher temperatures. Previously, BBN was considered the earliest direct probe of the Λ CDM model. With the advancement of sensitivities for GWBs at higher frequencies, we will therefore be able to directly probe the occurrence of anisotropic stress in the primordial plasma at much earlier times. We can thus rightfully claim: we are at the dawn of GW cosmology!

We live in the age of GW cosmology!

3 | PULSAR TIMING ARRAYS

Wissen ist Nacht!

— *Leitspruch der Dunkelheitsforschung,
Prof. Dr. Abdul Nachtigaller*

In June 2023, a significant milestone was achieved in the field of GW science and astrophysics: at an internationally acclaimed press conference the first evidence for the detection of a stochastic GW background in the nano-Hertz frequency band was unveiled. On the same day, multiple collaborations, including the North American NANOGrav [7], the Australian PPTA [8], the European EPTA in conjunction with the Indian InPTA [9], and the Chinese CPTA collaboration [10], released their latest data sets, all of which support this groundbreaking discovery.

In the first section of this chapter (section 3.1), we provide a brief overview of the history and current status of PTAs. Following that, in section 3.2, we explain how a monochromatic plane GW can manifest itself in the timing data of a single observed pulsar. In section 3.3, we express a stochastic GW signal as a superposition of plane waves, leading to the characteristic Hellings-Downs (HD) correlation function, for which PTAs have recently announced to have found evidence. Section 3.4 discusses the statistical methods used to separate a GWB from other noise sources in a PTA. A review of the latest results from the NANOGrav collaboration is presented in section 3.5. Finally, in section 3.6, this chapter concludes with a brief summary of the novel signal’s interpretation as an astrophysical background of supermassive black hole binaries (SMBHBs), being the strongest contender for alternative cosmological interpretations.

The nHz signal

Contents of this chapter

3.1 A SHORT HISTORY OF PULSAR TIMING

In 1967 Jocelyn Bell, then a PhD student under Antony Hewish, made a groundbreaking discovery by identifying the first pulsar J1921+2153 [93]. She observed radio sources with a very regular period of 1.337 s. At the time, such regular fluctuations in radio signals were often attributed to human activity rather than astronomical sources. Within the following years of her discovery, more pulsars were identified thanks to rapid advancements in radio astronomy. In particular, also the aforementioned Hulse-Taylor binary pulsar J1915+1606 which led to the first indirect

The first pulsar

discovery of GWs [81], rewarded with a Nobel Prize in 1993, was observed only a few years after Bell’s initial discovery.

Pulsars are spinning neutron stars

Already in the early 1930s, three decades before Bell’s discovery, the existence of neutron stars had been hypothesized by Landau, Baade, and Zwicky [94, 95]. Just a few months before Bell’s discovery, Pacini and Gold proposed the existence of rapidly rotating neutron stars which would emit pulses of electromagnetic radiation [96, 97]. The origin of their fast rotations is now understood to be due to the conservation of angular momentum from their progenitor stars, which become more compact and faster-spinning as the nuclear fusion within them ceases [98].

The lighthouse model

Neutron stars are sources of strong dipolar magnetic fields of up to 10^{11} T, whose axis is not necessarily aligned with the star’s axis of rotation. The neutron star can thus be modeled as a precessing magnetic dipole, which in turn generates an electric field. The latter accelerates charged particles within the neutron star’s magnetosphere, exciting beams of electromagnetic radiation in the radio band along the magnetic dipole’s direction, effectively turning the neutron star into a lighthouse sweeping radio beams through its host galaxy [99].

Millisecond pulsars

Typical pulsars have an orbital period of $P \simeq 0.1 - 1$ s and a spin-down rate of $\dot{P} \simeq 10^{-15}$, meaning their periods increase only by about 1 ns per month due to their high moment of inertia. Additionally, there is a population of millisecond pulsars which rotate much faster ($P \simeq 10$ ms), resulting in an even more stable orbital period ($\dot{P} \simeq 10^{-20}$). These millisecond pulsars are typically formed in binary systems where a regular pulsar accretes mass from its companion, thus gaining angular momentum from the binary system [98].

Hellings and Downs

The PTA idea

Due to the stability of their orbital periods, the sequence of pulses from millisecond pulsars can be predicted with extreme precision using a timing model. These models account for factors like spin-down \dot{P} and higher time-derivatives of the period P , the effects of the ionized interstellar medium, Earth’s proper motion, and possible delays from the orbital motion of binary pulsars. Importantly, GWs also affect the arrival times of these pulses, a phenomenon known since the late 1970s [100, 101]. The work of Hellings and Downs in the early 1980s provided a method to distinguish the effect of a GW background on arrival times from other noise sources [102]: By correlating timing residuals between several pulsars, a characteristic pattern due to the quadrupolar nature of gravitational radiation was predicted to be observable in case a GWB alters the time intervals between pulses.

This led to the concept of PTAs in the late 1980s [103], which promised to be able to disentangle the effects of clock errors (monopolar correlation) [104], a mismodeling of the Solar System’s barycenter (dipolar correlation) [105], and the searched-for HD correlation of GWs. Today, several PTA collaborations exist, namely NANOGrav, PPTA, EPTA, InPTA, and CPTA. Each collaboration observes a set of pulsars using radio telescopes, sometimes in various radio frequency bands. These col-

laborations (except for CPTA which acts as an observer) share their data under the International Pulsar Timing Array (IPTA) framework, thus drastically increasing the number of possible pulsar cross-correlations.

The lowest GW frequency a PTA can detect is determined by the inverse time span of the arrival data, $f \gtrsim 1/T$. For an indicative observational period of $T = 10$ yr, the lowest detectable frequency is slightly below 3 nHz. Lower-frequency GWs cannot be detected through pulsar timing variations but rather manifest themselves as contributions to the pulsar spin-down rate \dot{P} and higher derivatives of the period P . Thus, PTAs are generally not sensitive to lower-frequency GWs. The upper frequency limit is determined by how often a pulsar is observed, typically every few weeks or even once per week in the case of EPTA. The highest observable GW frequency is thus around $50 \text{ yr}^{-1} \simeq 1 \mu\text{Hz}$. However, sensitivity is lower at these frequencies due to noise-dominated timing residuals, i.e. the difference between the timing model and the actually observed times of arrival (TOAs) of pulses.

In their latest data release, the NANOGrav collaboration found evidence for a GW background with an amplitude of around $h^2 \Omega_{\text{gw}} \simeq 10^{-10} - 10^{-8}$ (see the previous fig. 2.8). It is not yet clear whether the signal is of astrophysical or cosmological origin. The prevailing null hypothesis for the signal's origin is an astrophysical background of SMBHBs [15]. However, due to the unexpectedly high amplitude and low spectral index of the measured signal, as well as unresolved issues like the final parsec problem (see sec. 3.6.1), this interpretation has been questioned [106].

nHz to μHz

An unexpectedly strong nHz background

3.2 GRAVITATIONAL WAVE EFFECTS ON A SINGLE PULSAR'S TIMING

We will now study the effect of GWs on a pulsar's emitted radio pulses. Before we study the effect of a stochastic background on a pulsar timing array, let us first consider the effect of a monochromatic and plane GW on a single pulsar's period. We will identify the coordinate origin with the barycenter of the Solar System (often referred to as *Earth* in the pulsar timing literature for simplicity) and refer to the pulsar position with $\mathbf{x} = \mathbf{n}_a$. The label a will allow us to generalize and correlate between two pulsars in the following section. We further work in TT gauge, such that the coordinate distances d_a between Earth and pulsars remain constant, whereas the proper distance between them oscillates. In that sense the following calculation is very similar to the one performed in section 2.4.1 and analogous to studying the response of a single interferometer arm to a GW. Both this and the following subsection are close to the calculations presented in [37, 107].

For a radio pulse moving in x -direction, the line element reads

$$0 = -dt^2 + (\delta_{ij} + h_{ij}^{\text{TT}}) dx^i dx^j = -dt^2 + (1 + h_{xx}^{\text{TT}}) dx^2. \quad (3.1)$$

We work in TT gauge

A pulse moving along the x -axis

The pulse moves from \mathbf{x} to the coordinate origin, so ignoring the effect of the metric perturbation we can write $\mathbf{x}(t) \simeq (t_{\text{obs}} - t)\hat{\mathbf{x}}$. To first order in the metric perturbation we obtain

$$dx = -\sqrt{\frac{dt^2}{1 + h_{xx}^{\text{TT}}(t, \mathbf{x}(t))}} = \left(1 - \frac{h_{xx}^{\text{TT}}(t, \mathbf{x}(t))}{2}\right) dt. \quad (3.2)$$

By integrating this expression from the emission of the radio pulse at t_{em} to the time of observation $t_{\text{obs}} \simeq t_{\text{em}} + d_a$, ignoring the effect of the metric perturbation in the integral boundary, we can obtain t_{obs} to first order in the metric perturbation,

$$t_{\text{obs}} = t_{\text{em}} + d_a + \frac{1}{2} \int_{t_{\text{em}}}^{t_{\text{obs}}} dt' h_{xx}^{\text{TT}}(t', \mathbf{x}(t')). \quad (3.3)$$

A pulse moving in direction $\hat{\mathbf{n}}_a$
The delay ΔP_a of a single pulse

We can easily generalize this expression to the case of a radio pulse in direction $\hat{\mathbf{n}}_a$ not necessarily coinciding with the x -axis by replacing $h_{xx}^{\text{TT}} \rightarrow n_a^i n_a^j h_{ij}^{\text{TT}}$, yielding

$$t_{\text{obs}} = t_{\text{em}} + d_a + \frac{n_a^i n_a^j}{2} \int_{t_{\text{em}}}^{t_{\text{obs}}} dt' h_{ij}^{\text{TT}}(t', (t_{\text{em}} + d_a - t') \hat{\mathbf{n}}_a). \quad (3.4)$$

The difference between t_{obs} and the observation time t'_{obs} of another radio pulse emitted at $t'_{\text{em}} = t_{\text{em}} + P_a$ hence reads $t'_{\text{obs}} - t_{\text{obs}} = P_a + \Delta P_a$ with

$$\Delta P_a = \frac{n_a^i n_a^j}{2} \int_{t_{\text{em}}}^{t_{\text{obs}}} dt' [h_{ij}^{\text{TT}}(t' + P_a, \mathbf{x}_a(t')) - h_{ij}^{\text{TT}}(t', \mathbf{x}_a(t'))], \quad (3.5)$$

where $\mathbf{x}_a(t') = (t_{\text{em}} + d_a - t') \hat{\mathbf{n}}_a$. Since the GW frequency is $f \simeq \mathcal{O}(\text{nHz})$, while the pulsar's rotational period is at the scale of $P_a \simeq \mathcal{O}(\text{ms})$, their product $fP_a \simeq 10^{-9}$ is tiny. Therefore, it is enough to expand the metric perturbation around t' to first order in fP_a to find an expression for the relative delay

$$\frac{\Delta P_a}{P_a} = \frac{n_a^i n_a^j}{2} \int_{t_{\text{em}}}^{t_{\text{obs}}} dt' \left[\frac{\partial}{\partial t'} h_{ij}^{\text{TT}}(t', \mathbf{x}) \right]_{\mathbf{x}=\mathbf{x}_a(t')} \quad (3.6)$$

between two pulses. A GW $h_{ij}^{\text{TT}}(t, \mathbf{x}_a) = \mathcal{A}_{ij}(\hat{\mathbf{n}}) \cos(2\pi f(t - \hat{\mathbf{n}} \cdot \hat{\mathbf{x}}_a))$ propagating in $\hat{\mathbf{n}}$ -direction hence induces a (sometimes referred to as *red-*) shift $z_a \equiv \frac{\Delta P_a}{P_a}$ of the pulse's TOA reading

$$z_a = \frac{n_a^i n_a^j \mathcal{A}_{ij}(\hat{\mathbf{n}})}{2(1 + \hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_a)} [\cos(2\pi f t_{\text{obs}}) - \cos(2\pi f t_{\text{em}} - 2\pi f \tau_a \hat{\mathbf{n}}_a \cdot \hat{\mathbf{n}})] \quad (3.7)$$

$$= \frac{n_a^i n_a^j}{2(1 + \hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_a)} [h_{ij}^{\text{TT}}(t_{\text{obs}}, \mathbf{x} = 0) - h_{ij}^{\text{TT}}(t_{\text{obs}} - \tau_a, \mathbf{x}_a)] \quad (3.8)$$

with $\tau_a = t_{\text{obs}} - t_{\text{em}} \sim \text{kpc} \sim 10^3 \text{ yr}$ being the runtime of the radio signal. The first term is usually referred to as the Earth term, as it only

depends on the local metric perturbation at the time of observation. In practice, the coordinate origin is however set to the barycenter of the Solar System, from which the time delay z_a at Earth can be inferred by translating into a rotating reference frame in the deterministic timing model discussed in section 3.4. This is done by adding up the effects of a Rømer time delay due to the annual motion of the Earth around the Sun, a Shapiro time delay due to the Solar System's own gravitational field, and an Einstein time delay due to the detectors proper motion around the Sun and the difference between coordinate time and proper time on Earth [107]. The second term in eq. (3.8), referred to as the pulsar term, depends on the metric perturbation at the location of the pulsar at the time of the pulse's emission, several thousand years ago.

*The “Earth” term
and the pulsar
term*

3.3 EFFECT OF A GWB ON AN ARRAY OF PULSARS

We can now apply eq. (3.8) to the case of a stochastic GWB. Using the plane wave expansion (2.65) and replacing t_{obs} for t , we immediately obtain

$$z_a(t) = \sum_{A=+,\times} \int_{-\infty}^{\infty} df \int d^2\hat{\mathbf{n}} \tilde{h}_A(f, \hat{\mathbf{n}}) F_a^A(\hat{\mathbf{n}}) e^{-2\pi ift} \\ \times \left(1 - e^{2\pi if\tau_a(1+\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_a)} \right) \quad (3.9)$$

*Expanding in plane
waves*

with the antenna response pattern function for pulsar a

$$F_a^A(\hat{\mathbf{n}}) \equiv \frac{n_a^i n_a^j e_{ij}^A(\hat{\mathbf{n}})}{2(1 + \hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_a)}. \quad (3.10)$$

As discussed previously, a Gaussian stochastic GWB is only characterized through its variance and mean, the latter of which we deliberately set to zero. Hence we need to compute the correlator $\langle z_a(t) z_b(t) \rangle$ to make progress, where the brackets refer to an ensemble average over realizations of the stochastic variable $\tilde{h}_A(f, \hat{\mathbf{n}})$. Plugging in the previous parameterization of the variance of a stationary, isotropic and unpolarized GW background from eq. (2.66), we obtain the correlator

$$\langle z_a(t) z_b(t) \rangle = \int_0^{\infty} df S_h(f) \int \frac{d^2\hat{\mathbf{n}}}{4\pi} \mathcal{K}_{ab}(f; \hat{\mathbf{n}}) \sum_{A=+,\times} F_a^A(\hat{\mathbf{n}}) F_b^A(\hat{\mathbf{n}}) \quad (3.11)$$

$$\text{with } \mathcal{K}_{ab}(f; \hat{\mathbf{n}}) = \left[1 - e^{-2\pi if\tau_a(1+\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_a)} \right] \left[1 - e^{2\pi if\tau_b(1+\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_b)} \right]. \quad (3.12)$$

*A GW background
is characterized by
its variance*

The function \mathcal{K}_{ab} is quickly oscillating with $f\tau_a \gtrsim 10$ for $f \gtrsim 1 \text{ nHz}$ and $\tau_a > 0.1 \text{ kpc}$ even for the closest observed pulsar. Unless $|1 + \hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_a| \lesssim 1/10$, i.e. unless the GW comes straight from the direction of the pulsar under consideration, we obtain $\mathcal{K}_{ab} \rightarrow 1$ for $a \neq b$. For $a = b$, one instead

$$\mathcal{K}_{ab} \rightarrow 1 + \delta_{ab}$$

obtains the auto-correlation limit $\mathcal{K}_{ab} \rightarrow 2$. Since $f\tau_a$ is large for most pulsars under consideration, we can hence safely set $\mathcal{K}_{ab}(f; \hat{\mathbf{n}}) \rightarrow 1 + \delta_{ab}$ and concentrate on the remaining integral over the product of detector pattern functions $F_a^A(\hat{\mathbf{n}})F_b^A(\hat{\mathbf{n}})$. This integral is elementary, but still requires a tedious calculation at whose end one arrives at the overlap reduction formula

$$\Gamma_{ab} \equiv \frac{3}{2} \int \frac{d^2 \hat{\mathbf{n}}}{4\pi} \mathcal{K}_{ab}(f; \hat{\mathbf{n}}) \sum_{A=+, \times} F_a^A(\hat{\mathbf{n}})F_b^A(\hat{\mathbf{n}}) \quad (3.13)$$

$$= \frac{3}{2} x_{ab} \ln x_{ab} - \frac{1}{4} x_{ab} + \frac{1}{2} \delta_{ab} + \frac{1}{2} \quad (3.14)$$

The HD curve
with $x_{ab} = \frac{1}{2}(1 - \cos \theta_{ab})$, where θ_{ab} is the angle between pulsars a and b as seen from Earth. The function $\Gamma(x_{ab})$ was first derived by Hellings and Downs [102] (for $a \neq b$) and is since known as the HD curve. The prefactor $\frac{3}{2}$ was introduced to ensure the normalization to $\Gamma_{ab} = 1$ for $a = b$.

Interpreting the HD curve

In fig. 3.1 the HD curve is plotted in comparison to a monopolar and a dipolar correlation function. A flat monopolar correlation would show up in the data analysis of a PTA in the case of clock errors which affect all pulse timestamps in the same way, corresponding to a maximal correlation between pulsars regardless of the specific pulsar pair. A dipolar correlation would instead indicate a mismodeling of the barycenter of the Solar System, leading to a secular motion of Earth with respect to the pulsars. Note that the HD curve is not a pure quadrupolar correlation function, as can be seen by it reaching 0.25 instead of 0.5 for two pulsars in opposite directions. Using a decomposition into Legendre polynomials one can however show that the HD curve can be well approximated as the sum of a quadrupolar and an octupolar term [107]. The reason for the occurrence of the octupole and higher-order multipolar correlations is that the denominator in the antenna response pattern (3.10) implies that there is a preferred direction $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_a = -1$, in which the response of the detector system of Earth and pulsar is largest. Note also that the maximum of the HD curve is reached only in the case of an auto-correlation $a = b$. Even pulsar pairs with a low angular separation can have a maximum cross-correlation of 0.5, because the product of two pulsar terms in \mathcal{K}_{ab} decorrelates quickly for pulsars with a separation larger than a GW wavelength [108].

For the correlator of shifts of pulse arrival rates we hence obtain

$$\langle z_a(t)z_b(t) \rangle = \frac{2}{3} \Gamma_{ab} \int_0^\infty df S_h(f). \quad (3.15)$$

The timing residual R_a and its variance r_{ab}

However, a PTA is not able to directly measure the shift z_a of a single pulse. Instead, it measures pulse arrival times, which shift from their regular pattern due to the accumulated timing residual

$$R_a(t) \equiv \int_0^t dt' z_a(t'). \quad (3.16)$$

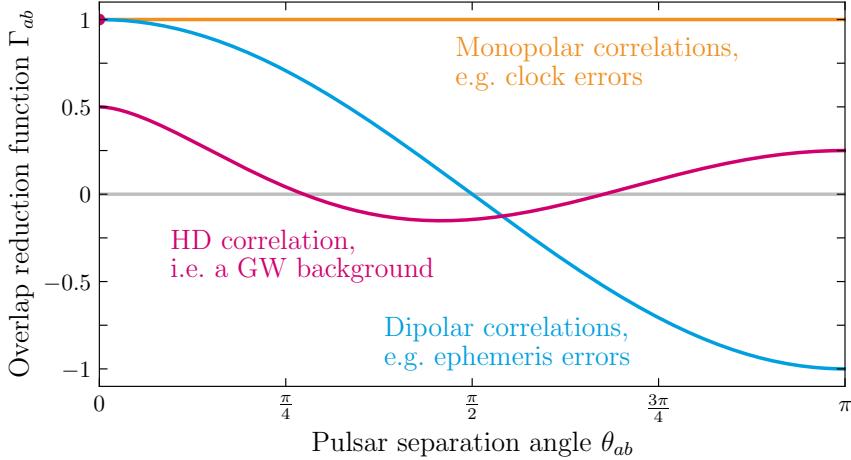


Figure 3.1: Comparison of the HD curve from eq. (3.14), a monopolar, and a dipolar correlation function. The latter two would for instance be expected if clock errors or Solar System ephemeris errors affected the TOAs instead of a GW background.

The quantity of interest to connect the previous considerations with actual PTA measurements is hence $r_{ab}(t) = \langle R_a(t)R_b(t) \rangle$. Plugging in the definition of $R_a(t)$ and eq. (3.9) for $z_a(t)$ we obtain the variance

$$r_{ab}(t) = \frac{2}{3}\Gamma_{ab} \int_0^\infty df \frac{S_h(f)}{(2\pi f)^2} 2(1 - \cos(2\pi ft)). \quad (3.17)$$

The time dependence occurs here due to the arbitrary lower integration boundary in the above definition of $R_a(t)$. Considering that the PTA only is sensitive to GWs in a frequency band in which $2(1 - \cos(2\pi ft)) = \mathcal{O}(1)$, one obtains [37]

$$r_{ab} \simeq \int_0^\infty df S_{ab}(f) \quad \text{with} \quad S_{ab}(f) \equiv \frac{2}{3}\Gamma_{ab} \frac{S_h(f)}{(2\pi f)^2}. \quad (3.18)$$

The function $S_{ab}(f)$ is referred to as the timing-residual cross-power spectral density and has units of s^3 . PTA collaborations often work with the parameterization $\Phi_{ab}(f) = S_{ab}(f)/T$ (with units of s^2) when doing their data analysis, where $T = \mathcal{O}(10 \text{ yr})$ is the total data set time span. The effect of the overlap reduction function can easily be factored out of this expression, $\Phi_{ab}(f) = \Gamma_{ab}\Phi(f)$, due to the previously identified approximate frequency-independence of \mathcal{K}_{ab} . The remaining spectrum $\Phi(f)$ (not carrying pulsar indices) only captures the effect of a given GW mode with frequency f on the variance on the timing residuals. The quantity $\rho(f) = \sqrt{\Phi(f)}$ in units of time is often referred to as the excess timing delay.

The timing-residual cross-power spectral density $S_{ab}(f)$

Expressing the GW spectrum through its characteristic strain amplitude $h_c(f)$ or its energy density spectrum $h^2\Omega_{\text{gw}}(f)$, see eq. (2.71), we end up with the following expressions for the excess timing delay

$$\rho(f) = \sqrt{\frac{S_h(f)}{6\pi^2 f^2 T}} = \sqrt{\frac{h_c^2(f)}{12\pi^2 f^3 T}} = \sqrt{\frac{H_{100}^2}{8\pi^4 f^5 T} h^2\Omega_{\text{gw}}(f)}. \quad (3.19)$$

The excess timing delay $\rho(f)$ and its relation to $\Omega_{\text{gw}}(f)$

Generalizing the overlap reduction formula

Given a measurement of $\rho(f)$, we can hence easily infer the underlying GW spectrum. Typical values of $\rho(f)$ that PTAs quote are between 1 ns and 1 μ s for frequencies in the range 1 – 30 nHz (see for instance fig. 3.2 in the following section).

Note that in the above calculation we assumed that the stochastic GWB under consideration is isotropic. In case the GWB is instead of astrophysical origin, this assumption cannot be sustained and a more generalized overlap reduction formula needs to be used. In the presence of anisotropies, one finds that the overlap reduction formula becomes a function of the individual pulsars' positions on the celestial sphere [109]. Further, GR modifications which lead to GWs with more than two propagating tensorial metric components lead to a dependence of the overlap reduction formula on the pulsar distances and the GW frequency, rendering searches for the HD correlation and deviations from it a powerful probe of new physics in their own right [110]. Within this thesis, we will only consider the HD curve as a possible overlap reduction formula or set $\Gamma_{ab} = \delta_{ab}$ for computational reasons to be explained in the next section, thus ignoring the effect of the GW polarization on the pulse arrival times, when trying to fit them to a specific backgrounds $h^2\Omega_{\text{gw}}(f)$ in chapters 5 and 7.

3.4 THE PTA LIKELIHOOD

Separating deterministic signals from interesting noise

In the above derivation we assumed that only GWs lead to a change of the arrival times of a given pulsar's radio pulses. This is obviously an oversimplification. In practice, the mechanisms which predict the TOA of a pulse are a combination of both deterministic *and* stochastic processes, the latter of which include the effect of GWs,¹

$$t_{ai}^{\text{TOA}} = t_{ai}^{\text{det}} + \delta t_{ai}^{\text{stoch}}. \quad (3.20)$$

The index i here indicates the i -th pulse of pulsar a with respect to an arbitrary initial point in time at which data taking starts. The goal of this section will now be to identify how a GW-induced timing residual $R_a(t_i)$ (defined in eq. (3.16)) contributes to the stochastic part $\delta t_{ai}^{\text{stoch}}$ of the TOA. Eventually we will present how one can infer a GW spectrum

¹ We will ignore the effect of continuous GWs (e.g. from a single close-by SMBHB inspiral) for brevity, which would show up as an unmodeled contribution to the deterministic timing model. So far, no significant evidence for such continuous GWs has been found with PTAs [15].

$h^2\Omega_{\text{gw}}(f)$ from pulsar timing data. To do so, we will summarize the analysis pipeline used by the NANOGrav collaboration, described in detail in ref. [107].

In a first step, a deterministic timing model needs to be found for each pulsar of the PTA. To do so, there exist the highly specialized codes **TEMPO**, **TEMPO2** [111–113] and **PINT** [114]. They allow to perform a fit of the observed TOAs t_{ai}^{TOA} to a timing model (also referred to as timing ephemeris) $t_{ai}^{\text{det}}(\beta_a)$, which depends on a set of deterministic timing parameters β_a , including the period of the pulsar, its spindown, possible pulsar glitches, the aforementioned Solar System ephemeris, the precise position of the pulsar in the celestial sphere and its motion therein, as well as additional Keplerian parameters if the pulsar is part of a binary system. After performing a least-squares fit, one obtains the best-fit vector β_a^0 and the corresponding timing residuals

$$\delta t_{ai} = t_{ai}^{\text{TOA}} - t_{ai}^{\text{det}}(\beta_a^0) \quad (3.21)$$

for each pulsar.

No matter how accurate the first step is done, the process of finding a timing model is limited by it only including deterministic processes. The found timing model will hence be very good, but not a perfect fit for the observed TOAs t_{ai}^{TOA} . Denoting the true (and unknown) timing ephemeris with β_a^{true} and interpreting the TOA index i as the component of a vector, we can write eq. (3.20) as $\mathbf{t}_a^{\text{TOA}} = \mathbf{t}_a^{\text{det}}(\beta_a^{\text{true}}) + \boldsymbol{\delta t}_a^{\text{stoch}}$. Further, defining the difference between our initial timing model and the true model as $\epsilon_a = \beta_a^{\text{true}} - \beta_a^0$ and plugging into eq. (3.21) we find

$$\begin{aligned} \boldsymbol{\delta t}_a &= \mathbf{t}_a^{\text{det}}(\beta_a^0 + \epsilon_a) - \mathbf{t}_a^{\text{det}}(\beta_a^0) + \boldsymbol{\delta t}_a^{\text{stoch}} \\ &= M_a \epsilon_a + \boldsymbol{\delta t}_a^{\text{stoch}} + \mathcal{O}(\epsilon_a^2) \quad \text{with} \quad M_a = \nabla_{\beta_a} \mathbf{t}_a|_{\beta_a^0}. \end{aligned} \quad (3.22)$$

The design matrix M_a has the shape $(N_{\text{TOA}} \times m)$, where N_{TOA} is the length of the vector \mathbf{t}_a corresponding to the number of TOAs for pulsar a , and m refers to the number of timing ephemeris parameters in β_a .

The goal is now to identify the different noise sources which contribute as random Gaussian processes in $\boldsymbol{\delta t}_a^{\text{stoch}}$, including a GWB. Typically, one distinguishes between red and white noise sources,

$$\boldsymbol{\delta t}_a^{\text{stoch}} = \boldsymbol{\delta t}_a^{\text{red}} + \boldsymbol{\delta t}_a^{\text{white}}. \quad (3.23)$$

White noise is independent of the Fourier modes of $\boldsymbol{\delta t}_a^{\text{stoch}}$, whereas red noise instead has a frequency dependence and amplitude that is largest at the lowest frequencies. Red noise is therefore sourced by processes which operate on long timescales of the order of months to decades.

The white noise $\boldsymbol{\delta t}_a^{\text{white}}$ has different sources connected to radiometer noise, instrumental effects, and pulsar phase jitter. Often, white noise sources are also referred to by their technical names **EFAC**, **EQUAD** and **ECORR**. The specific origins of these various noises are not central to the focus of this thesis. Let us only note here, that they are related

Step 1: The timing model

*Step 2:
Understanding the timing residuals*

Stochastic noise sources are red or white

White noises

to the specific combinations of radio telescope receiver systems and data-processing backends. The data analysis is divided into epochs of 20 – 30 minutes, within which a pulse profile template is generated by folding over many observed realizations of a given pulsar’s signal in that epoch. The pulse template generation procedure adds jitter noise to the TOAs, which are retrieved by identifying pulses in shape of the template within the data. It therefore makes sense to distinguish the above noise sources acting on different time-scales. For instance, the **ECORR** noise is uncorrelated between two epochs but fully correlated for different radio frequency channels through which a pulsar is observed.

Red noises

For the red noise one decomposes the timing residual into Fourier modes which are multiples of the lowest possible frequency T^{-1} which can be probed by the PTA,

$$\delta t_{ai}^{\text{red}} = \sum_{k=1}^{N_f} \left[a_{ak} \sin\left(\frac{2\pi k t_{ai}^{\text{TOA}}}{T}\right) + b_{ak} \cos\left(\frac{2\pi k t_{ai}^{\text{TOA}}}{T}\right) \right]. \quad (3.24)$$

Usually, only the lowest $N_f = 50$ Fourier modes are used in the analysis, as white noise dominates at high frequencies. Again, it is possible to write this more compactly as a vector

$$\boldsymbol{\delta t}_a^{\text{red}} = F_a \mathbf{c}_a, \quad \text{where } \mathbf{c}_a = (a_{a1}, b_{a1}, \dots, a_{aN_f}, b_{aN_f})^T \quad (3.25)$$

and F_a is a matrix of shape $(N_{\text{TOA}} \times N_f)$ containing alternating sine and cosine terms in each row with increasing arguments $\frac{2\pi k}{T} t_{ai}^{\text{TOA}}$ in the k -th column and i -th row. The vector \mathbf{c}_a hence includes both the amplitudes for pulsar-intrinsic red noise of pulsar a and the GW spectrum, present in all pulsars’ TOAs.

How to distinguish GWs from other noise

To distinguish the stochastic red noise sourced by a GWB from pulsar-intrinsic red noise, we need to correlate the timing residuals between pulsars. We assume that the Gaussian process underlying the stochastic part of the TOAs has mean zero, as any residual mean would have been absorbed by the deterministic timing model already. The full statistical information of the distributions underlying the Gaussian processes is therefore contained in the large cross-correlation matrix

$$C_{(ai)(bj)} = \langle \delta t_{ai} \delta t_{bj} \rangle \quad (3.26)$$

between the i -th TOA of pulsar a and j -th TOA of pulsar b . As discussed above, this matrix splits into the different contributions $C = C^{\text{GWB}} + C^{\text{IRN}} + C^{\text{WN}}$ from a GWB, pulsar-intrinsic red noise and white noise. The different components can be expressed as

$$C_{(ai)(bj)}^{\text{GWB}} = \Gamma_{ab} \Phi_i \delta_{ij}, \quad (3.27a)$$

$$C_{(ai)(bj)}^{\text{IRN}} = \delta_{ab} \kappa_{ai} \delta_{ij}, \quad (3.27b)$$

$$\text{and } C_{(ai)(bj)}^{\text{WN}} = [F_a^2 \delta_{ij} + J_a^2 \delta_{e(i)e(j)}] \delta_{ab}, \quad (3.27c)$$

where Γ_{ab} is the HD-correlation from eq. (3.14) and Φ_i is related to the GW spectrum $\Phi(f)$ (see eq. (3.19)) evaluated at the i -th Fourier

mode. Similarly, κ_{ai} is the intrinsic red noise power spectrum of pulsar a . The factors F_a and J_a refer to **EFAC + EQUAD** and **ECORR** white noise respectively, where the latter is only correlated when TOA i lies in the same epoch $e(i) = e(j)$ as TOA j .

In addition to the aforementioned noise sources, variations in the refractive index of the interstellar medium also contribute significantly to the timing residuals. These variations arise from (*i*) deterministic, time-dependent changes in the line-of-sight from Earth to the pulsar, such as the pulsar's peculiar motion, and (*ii*) stochastic factors, such as the distribution of free electrons in the interstellar medium. Since the electron number density behaves more like a stochastic process than a deterministically modeled one, it is reasonable to include an additional pulsar-intrinsic red noise component with a distinct spectrum to account for this noise source. Modeling dispersion measure variations as a purely deterministic effect in the timing ephemeris β_a is referred to as the DMX model. If dispersion measure variations are instead modeled as a Gaussian process, they show up as an additional contribution in $C_{(ai)(bj)}$ usually abbreviated as DMGP. As dispersion measure variations act on the same long timescales as GWs with nHz frequencies it further makes sense to specify whether a GWB spectrum was inferred using the DMX or DMGP model. The analysis of the latest NANOGrav 15yr data set (see ref. [7] and fig. 5 therein) indicates, that the DMGP model yields slightly smaller GWB amplitudes, but comparable evidence for an HD-correlation of the signal. Within this thesis, we will only make use of the DMX model in our PTA analysis in chapters 4 and 7.

*A note on
dispersion measure
variations*

The full PTA likelihood for the set of all timing residuals $\{\delta t_a\}$ for all pulsars $a = 1, \dots, N_p$ is hence given by a multidimensional Gaussian with zero mean and the cross-correlation $C_{(ai)(bj)}$ [37],

$$\mathcal{L}_{\text{PTA}}(\{\delta t_a\}|\boldsymbol{\theta}) = \frac{\exp\left[-\frac{1}{2}\sum_{(ai),(bj)} \delta t_{ai} C_{(ai)(bj)}^{-1} \delta t_{bj}\right]}{\sqrt{\det(2\pi C)}}. \quad (3.28)$$

The PTA likelihood

The parameter vector $\boldsymbol{\theta}$ contains the timing ephemeris shifts ϵ_a , the red noise Fourier amplitudes \mathbf{c}_a , the **EFAC + EQUAD** and **ECORR** parameters F_a and J_a for all N_p pulsars, as well as the hyper-parameters that go into the parameterization of the GWB spectrum $h^2\Omega_{\text{gw}}(f)$ and the N_p pulsar-intrinsic red noise spectra $\kappa_a(f)$. In a typical PTA data analysis, both $\Phi(f)$ and $\kappa_a(f)$ are assumed to follow power-laws

$$\Phi(f) = \frac{h_c^2(f)}{12\pi^2 f^3 T} \frac{1}{T} = \frac{A^2}{12\pi^2} \frac{1}{T} \left(\frac{f}{1 \text{ yr}^{-1}}\right)^{-\gamma} \text{yr}^2 \quad (3.29a)$$

$$\kappa_a(f) = \frac{A_a^2}{12\pi^2} \frac{1}{T} \left(\frac{f}{1 \text{ yr}^{-1}}\right)^{-\gamma_a} \text{yr}^2 \quad (3.29b)$$

*Power-law red
noise spectra*

with a spectral tilt $\gamma = 2 - 2\alpha$ for $h_c(f) = A(f/1\text{yr}^{-1})^\alpha$. This way, one obtains $2 + 2N_p$ hyper-parameters. Another way of parameterizing the GW spectrum is the so-called free-spectral analysis, in which $\Phi(f)$ is instead decomposed into individual Fourier modes whose amplitudes act

Hyper-parameters

as to-be-inferred hyper-parameters. In chapter 5 we will instead treat phase transition parameters like α , β/H and T_{perc} as hyper-parameters; in chapter 7 the PBH parameters f_{PBH} , m_{PBH} and δ_{dc} are used as hyper-parameters.

enterprise

The sum in the exponent of \mathcal{L}_{PTA} runs over all combinations of all TOAs of all pulsars. The matrix C is hence a high-dimensional, dense, quadratic matrix consisting of $(N_p \times N_p)$ blocks on the diagonal δ_{ab} , each of which contains GW, as well as red and white noise contributions. The off-diagonal parts of the matrix contain the pulsar cross-correlations. While it is computationally feasible to evaluate this likelihood for a PTA consisting of only a few pulsars, it certainly is not possible for a full PTA with many pulsars due to the required inversion C^{-1} in the exponent. Luckily, one is usually only interested in the hyper-parameters mentioned above. As the full likelihood is Gaussian, one can analytically marginalize over the nuisance parameters ϵ_a and c_a of all pulsars. The remaining marginalized PTA likelihood is written out explicitly in ref. [107] and implemented in the code `enterprise` [115, 116], which will be used to derive the results of chapters 5 and 7.

MCMC chains

The remaining parameter space of at least $2 + 2N_p$ hyper-parameters is still huge for today’s PTAs with $N_p = \mathcal{O}(50 - 100)$. The parameter space is usually explored by running highly optimized MCMC codes (in particular `PTMCMC` [117]) to pull samples from the likelihood in order to produce chains of the hyper-parameters, from which one can then compute their posterior distributions as well as correlations between them using the methods of Bayesian inference. As we will discuss in chapter 5.C.1, this also allows Bayesian model comparisons between models with different sets of hyper-parameters.

*The need for speed:
`ceffyl`*

Recently, the code `ceffyl` was introduced by the NANOGrav collaboration, which allows a much faster hyper-parameter inference [118]. To do so, it was realized that at low signal-to-noise ratios the impact of the HD-correlation on the free-spectral analysis is negligible. One can hence artificially set the overlap reduction function $\Gamma_{ab} = \delta_{ab}$ in eq. (3.27a) to then perform a free-spectral analysis of the common uncorrelated red noise (CURN). The matrix $C_{(ai)(bj)}^{\text{GWB}}$ is hence effectively replaced by $C_{(ai)(bj)}^{\text{CURN}} = \delta_{ab}\Phi_i\delta_{ij}$, where Φ_i can now be understood as an estimator of the GWB spectrum. The resulting likelihood for the timing residuals given a spectrum $\rho(f) = \sqrt{\Phi(f)}$ then factorizes into the product of the individual pulsars’ contributions to the CURN. Assuming further that the Fourier modes ρ_k are statistically independent, one arrives at the simple product of $N_f^{\text{gw}} = \mathcal{O}(10)$ “violin” likelihoods

$$\mathcal{L}_{\text{ceffyl}}(\{\delta t_a\}|\boldsymbol{\eta}) = \prod_{k=1}^{N_f^{\text{gw}}} \mathcal{L}_k(\{\delta t_a\}|\rho_k(\boldsymbol{\eta})), \quad (3.30)$$

where $\mathcal{L}_k(\{\delta t_a\}|\rho_k(\boldsymbol{\eta}))$ is the likelihood to find a set of timing residuals $\{\delta t_a\}$ given a GWB amplitude ρ_k at Fourier mode k for a specific set of GW hyper-parameters $\boldsymbol{\eta}$. The likelihoods \mathcal{L}_k can be found through a

free-spectral analysis of the CURN and directly correspond to the violins in fig. 2.8. Pictorially speaking, `ceffyl` hence computes how well a given realization of the GWB signal described by η fits those violins. If the violins were perfectly Gaussian distributions, the latter likelihood would simplify to that of a least-squares fit. In turn, the `ceffyl` likelihood can also be thought of as the likelihood for a fit through data points whose error bars are indicated by the widths of the violins. The computation of the likelihood thus becomes a trivial task, which only requires a set of well-sampled Bayesian spectrograms (“violins”) for ρ_k . In fact, the hyper-parameter posteriors obtained from `ceffyl` and `enterprise` are indeed often indistinguishable (cf. fig. 3 of ref. [118]).

Learning how to play the violin

However, one should remain cautious of this quick method since it can lead to statistical fallacies like the following: Even though the PTA violins only span a certain range within the predefined prior ranges of the individual ρ_k , such that in particular $\mathcal{L}_k(\{\delta t_a\}|0) = 0$, this does not mean that the null-signal hypothesis could be rejected with arbitrary statistical significance. The distribution of ρ_k is a Bayesian posterior and should only be interpreted as such. Hence, one needs to be careful when interpreting the resulting hyper-parameter posteriors in regions of their parameter space where they describe a signal that goes above or below the limits of even a single violin for which $\mathcal{L}_k = 0$. This is precisely the reason why the full likelihood implemented in `enterprise` had to be used to derive the results in chapter 5 and was still required as a cross-check for the calculations presented in chapter 7.

A note on statistical fallacies

3.5 EVIDENCE FOR A GRAVITATIONAL WAVE BACKGROUND

In 2023, the NANOGrav collaboration made a groundbreaking announcement: compelling evidence for the detection of a stochastic signal that follows the HD-correlation was found [7]. This observation is widely regarded as the first detection of a GWB. While other collaborations such as EPTA [9], PPTA [8], and CPTA [10] reported similar results on the same day, this section specifically focuses on the findings of NANOGrav, as they were the only ones to make their full data set and analysis tools publicly available.

Multiple sources of evidence

NANOGrav’s analysis utilized data from 67 pulsars, each observed for more than three years. The total data span is $T = 16.03$ yr, corresponding to a lowest testable GW frequency of $f = 1.98$ nHz. The Bayes factor comparing a GWB modeled as a power-law to a model with only pulsar-intrinsic red noise exceeds 10^{14} . When comparing a power-law GW background to a CURN model, the Bayes factor ranged from 200 to 1000, depending on the number of frequency bins N_f^{gw} considered in the analysis. This Bayes factor translates to a false-alarm probability of $p = 5 \cdot 10^{-5}$ to 10^{-3} , corresponding to $3 - 4\sigma$ evidence, depending on

$3 - 4\sigma$ in favor of the HD correlation

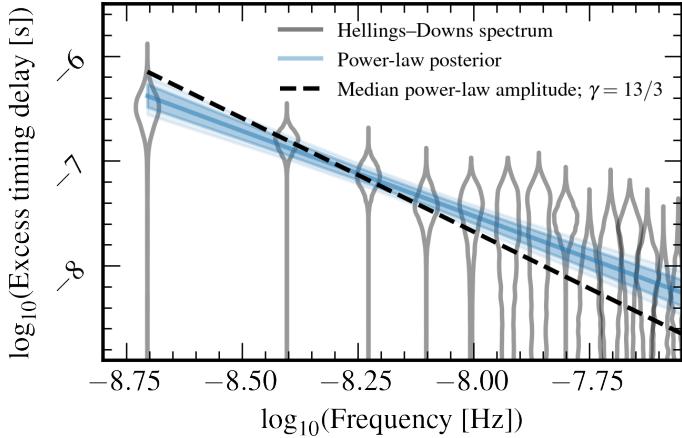


Figure 3.2: Results of a free-spectral analysis and fit to power-law GWB signals of NANOGrav’s 15-year data set. The gray violins represent GW amplitudes’ posteriors at different frequencies $f_k = k/T$. The dashed line indicates the best-fit GWB power-law model with a fixed spectral index $\gamma = 13/3$. The blue bands represent the 1σ and 2σ posterior intervals for the power-law model with a variable spectral index. This figure was taken from ref. [7].

the precise method of translation. Dipolar and monopolar correlations were ruled out with respect to CURN with Bayes factors below 10^{-7} and 10^{-8} , respectively [7].

$\gamma < 13/3$ is preferred

Fig. 3.2 presents the results of NANOGrav’s free-spectral analysis. The vertical axis shows the Bayesian posteriors of the Fourier amplitudes $\rho(f_k) = \sqrt{\Phi(f_k)}$ for an HD-correlated stochastic process at frequencies $f_k = k/T$, where T is the total data set time span. The dashed line represents the best-fit GWB of power-law shape with $\gamma = 13/3$ fixed. The blue regions indicate the 1 and 2σ posterior bands for a power-law with a variable spectral index γ . The tilt of the spectrum, showing most power in the lower frequency bins, illustrates why it is referred to as red noise. At high frequencies, the width of the posteriors $\rho(f_k)$ broadens due to higher levels of white noise contaminating the GWB signal. Comparing the blue bands to the black dashed line shows that the data prefers a spectral index below $\gamma = 13/3$.

Bayesian reconstruction of the HD curve

Fig. 3.3 showcases the Bayesian reconstruction of the overlap reduction function for the common red noise. This function was modeled as a cubic spline, while a power-law with a variable spectral index γ was used for the GWB. The posterior distributions of the overlap reduction function were computed at pre-selected positions, including the point of minimal HD correlation at 49 degrees, the zero-crossings of the HD curve, and the minimal and maximal angular separations. For reference, the HD curve is shown as a dotted line.

The robustness of these results was confirmed through various methods, including a separate frequentist analysis, a dropout analysis (leaving

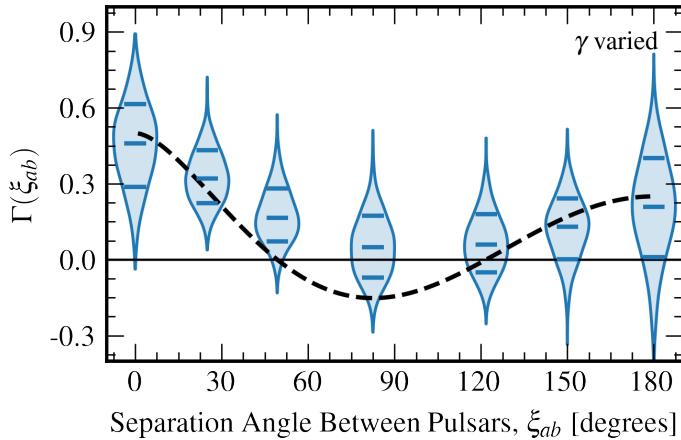


Figure 3.3: Bayesian reconstruction of the overlap reduction function for the common red noise signal in NANOGrav’s 15-year data set. The blue violins display the posterior distributions of the overlap reduction function at selected separation angles, with the dashed line representing the expected HD curve for a stochastic gravitational wave background. This figure was taken from ref. [7].

out some pulsars), and analyzing data from individual radio telescopes. While there is some dependence on the choice between the DMX and DMGP model for hyper-parameter inference, this does not affect the evidence for the HD correlation [7]. The evidence for the GWB is expected to increase further with time. The anticipated 5σ detection limit might already be achieved within the upcoming third IPTA data release, which will combine datasets from NANOGrav, EPTA, InPTA, and PPTA, encompassing observations of 80 pulsars over a 24-year data span.

Following the PTA announcements of 2023, a large number of articles studied a hypothetical cosmological origin of the novel signal. Possible origins of the PTA signal include the GWB emission through FOPTs [1, 119–122] or merging primordial black holes [3, 123], which we will investigate in depth in the following chapters of this thesis. Other hypothesized cosmological sources for the nano-Hertz signal include cosmic strings [124–128], domain walls [129], tensor modes produced in inflation [130–132], scalar-induced GWs [133], the production of primordial black holes [132, 134–137] and long-lived turbulence after a strong first-order QCD PT [138]. The arguably strongest contender of these explanations, each of which includes an extension of the SM in one or the other way, is a stochastic background emitted by a population of SMBHBs, being the subject of the following section.

*Robustness of
NANOGrav’s
findings*

*Cosmology or
astrophysics?*

3.6 SUPERMASSIVE BLACK HOLE BINARIES

*Astrophysical GW
foregrounds*

In any part of the spectrum of stochastic GWBs within the nHz to kHz range, one expects astrophysical foregrounds with amplitudes between $h^2\Omega_{\text{gw}} \simeq 10^{-12} - 10^{-9}$, limiting the sensitivity to primordial backgrounds (cf. fig. 4 of ref. [139] and references therein). In this section we will discuss the astrophysical GWB in the nHz range stemming from mergers of SMBHBs, i.e. black holes with $M = 10^8 - 10^{10} M_\odot$, relevant for PTA searches of primordial backgrounds.

*Supermassive black
holes*

There is general agreement that supermassive black holes reside at the centers of most large galaxies, observable primarily through their interactions with surrounding matter, for instance as quasars [140, 141]. Only in 2019, it was possible to for the first time directly observe the shadow of the supermassive black hole Sagittarius A* residing at the center of our own galaxy [142], after the existence of a black hole at the center of our galaxy with $M = 4 \cdot 10^9 M_\odot$ was predicted already in the late 1990s by the 2020 Nobel laureates Andrea Ghez and Reinhard Genzel [143, 144]. As the growth of cosmic structure occurs hierarchically, through mergers of smaller galaxies into larger ones, a series of black hole mergers is expected in the recent cosmic history, giving rise to a stochastic GWB.

*Open problems
concerning
SMBHBs*

The GWB signal from SMBHBs is still a subject of active research and several key processes have not been fully understood, yet: For instance, it is still unclear how a large population of supermassive black holes can form given the maximal speed of accretion set by the Eddington formula [30], a challenge known as the timing problem. Moreover, the merger process itself is complicated by the so-called final parsec problem [16, 24, 145]. While the existence of many SMBHBs with separations ranging from hundreds to thousands of parsecs have been observed in multi-messenger astronomy [146], the closest observed pair of supermassive black holes was still separated by roughly 7 pc [147]. For binaries to emit GWs in the nHz band, however, separations at the milliparsec-scale are required, which current optical telescopes are far from being able to resolve.

*Dynamical friction
(kpc – pc)*

3.6.1 Binary evolution and the final parsec problem

The main process assumed to dominate the hardening (i.e., the dissipation of binding energy) of a binary system with kiloparsec to parsec separations is dynamical friction. It is sourced by viscous drag forces of SMBHBs from many weak and long-range gravitational scatterings with smaller bodies. This inspiral down to a parsec between the black holes takes roughly $\mathcal{O}(\text{Gyr})$ times [16, 24]. At lower distances, dynamical friction ceases to be efficient because of the increased radial velocity of the binary SMBHBs.

Below parsec separations, a process known as stellar loss-cone scattering becomes relevant instead. In this process, individual stars close to the galactic core are scattered out of the central region due to a gravitational slingshot mechanism. The hardening of the binary hence continues until this reservoir of stars within the so-called loss-cone is depleted. The original definition of the final parsec problem was the stalling of the binary evolution at this stage. Today, there are several pieces of evidence supporting the claim that there is an ongoing supply of stars in the post-merger galaxy [21–23].

Stellar loss-cone scattering ($\lesssim 1\text{ pc}$)

At even lower separations of a hundredth to a thousandth of a parsec, viscous dissipation of binding energy to the gaseous circumbinary disk could become important. It is, however, not yet clear whether this process actually hardens the binary [17,18] or rather widens it [20], giving rise to a new final parsec problem. Another interesting mechanism, which however requires new physics, acting at the same distance scales is dynamical friction of a self-interacting DM fluid surrounding the SMBHB [148].

Viscous dissipation ($\lesssim 10\text{ mpc}$)

If a binary can pass this barrier and hardens down to below-milliparsec separations, the binary evolution decouples from astrophysical interactions and instead becomes driven by GW emission alone until the black holes coalesce at microparsec separations. It has been proposed that another mechanism might contribute to the hardening of binaries, through so-called triplet interactions with a close-by third supermassive black hole [19]. It appears reasonable that in the process of hierarchical structure formation, a third galaxy might come close to the binary whose orbit then becomes more eccentric, which accelerates its coalescence due to an increased GW emission. In fact, this will be the mechanism which will ensure a solution of the final parsec problem for the merging PBHs considered in chapter 7. To date the final parsec problem has not been solved on theoretical grounds. Given the recent advances of PTAs it is however conceivable that the underlying mechanism can be inferred through the emitted GW spectrum, which we will discuss in the the following section.

GW emission until coalescence (mpc – μpc)

3.6.2 GWBs from inspiraling binaries

Assuming that the final parsec problem can be resolved, the GWB from a population of SMBHBs could explain the measured nHz background [149]. We will now briefly review how a power-law spectrum can be derived from simple scaling relationships and where possible deviations from it can stem from. The introduced formalism will be used in chapter 7, in which we will test the assumption that the novel signal stems from inspiraling SMBHBs of primordial origin.

A simple power-law

Any isotropic background from a large enough compact binary population can be expressed as [37]

$$h^2 \Omega_{\text{gw}}(f) = \frac{h^2}{\rho_c^0} \frac{d\rho_{\text{gw}}}{d \log f} = \frac{h^2}{\rho_c^0} \int_0^\infty dz \frac{dn}{dz} \frac{1}{1+z} \frac{dE_{\text{gw}}^r}{d \log f_r} \Big|_{f_r=f(1+z)}, \quad (3.31)$$

Summing over sources and integrating over the expansion history

The quadrupole formula

$\omega_{\text{gw}} = 2\omega$

The luminosity in GWs

where dn/dz is the number density distribution of binaries over redshift z , $dE_{\text{gw}}^r/d \log f_r$ is the source-frame GW energy spectrum per logarithmic frequency interval, $f_r = f(1+z)$ is the GW frequency in the source frame, and the factor $1/(1+z)$ accounts for the redshift of the energy $E_{\text{gw}} = E_{\text{gw}}^r/(1+z)$ to today. The interpretation of the above equation is straightforward: to find the amplitude of the GW spectrum at a frequency f , we sum over all contributions from binaries emitting frequencies f_r corresponding to a frequency f today and then integrate over the cosmic expansion history.

To get a feeling for the quantities involved in eq. (3.31) let us first compute the expected frequency dependence of the spectrum dE_{gw}^r/df_r for a single binary. In the weak-field limit of small velocities and at large radial separations r from the binary, the GW equation of motion (2.37) can be solved by the famous quadrupole formula [14]

$$h_{ij}^{\text{TT}} = \frac{\Lambda_{ij,kl}}{4\pi m_{\text{Pl}}^2 r} \ddot{Q}_{kl} \left(t - \frac{r}{c} \right) \quad \text{with} \quad \Lambda_{ij,kl}(\hat{\mathbf{n}}) = P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl}$$

and
$$Q_{ij} = \int d^3x \rho(t, \mathbf{x}) \left(x_i x_j - \frac{1}{3}\delta_{ij}r^2 \right), \quad (3.32)$$

where $\Lambda_{ij,kl}(\hat{\mathbf{n}})$ is simply known as the Λ -tensor. It can be expressed through the tensor $P_{ij}(\hat{\mathbf{n}}) \equiv \delta_{ij} - n_i n_j$ and it is commonly used to project out the TT-part of tensors. A derivation of the above equation can for instance be found in chapter 3 of ref. [14]. Assuming that the binary consists of two compact objects of mass M on a stable circular orbit with radius R in the x - y plane and orbital frequency ω , one obtains

$$h_{ij}^{\text{TT}} = -\frac{MR^2\omega^2}{\pi m_{\text{Pl}}^2 r} \begin{pmatrix} \cos 2\omega t & \sin 2\omega t & 0 \\ \sin \omega t & -\cos 2\omega t & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.33)$$

Using the stress-energy tensor of a GW from eq. (2.50) one can obtain an expression for the energy loss of the binary through GW emission per unit solid angle. Integrating over a large enough sphere around the binary, one obtains the gravitational luminosity of the source

$$P_{\text{gw}} = \frac{dE_{\text{gw}}^r}{dt_r} = \frac{\langle \ddot{Q}_{ij} \ddot{Q}^{ij} \rangle}{40\pi m_{\text{Pl}}^2} \propto GM^2R^2\omega^2 \propto G^{7/3}M^{10/3}\omega^{10/3}, \quad (3.34)$$

where in the last step we used Kepler's law $\omega^2 = \frac{GM}{4R^3}$ to eliminate the R -dependence. Obviously, this radiative energy loss will reduce the

binding energy $E_b \propto GM^2/R$ of the binary, thus hardening the binary. Working in the limit that this radiative energy loss takes place on much longer timescales than the orbital motion, we can still treat the orbits as quasi-circular and equate $\dot{E}_b = P_{\text{gw}}$ in the cosmic rest-frame of the binary. Plugging in Kepler's law another time to eliminate R in favor of ω and using the previously derived expression for P_{gw} , one immediately arrives at $\dot{\omega} \propto G^{13/3} M^{17/3} \omega^{11/3}$.

Realizing that the GW frequency is proportional to the orbital frequency, $2\pi f_r = 2\omega$ (see eq. 3.33), we can hence perform the simple approximation for the spectrum emitted by a compact binary

$$\frac{dE_{\text{gw}}^r}{d \log f_r} = f_r \frac{dE_{\text{gw}}^r}{dt_r} \frac{dt_r}{df_r} \propto f_r \times f_r^{10/3} \times f_r^{-11/3} = f_r^{2/3}. \quad (3.35)$$

Since this expression carries the full frequency dependence of eq. (3.31), we find that $h^2 \Omega_{\text{gw}}(f) \propto f^{2/3}$ or equivalently $h_c \propto f^{-2/3}$ and $\Phi(f) \propto f^{-13/3}$ (cf. eq. (3.29a)). The exponents of $h_c(f)$ and $\Phi(f)$ are usually referred to as $\alpha = -2/3$ and $\gamma = 2 - 2\alpha = 13/3$, respectively. In fact, the proportionalities derived above also hold for a general circular orbit of compact binaries with different masses, which was first appreciated by Phinney in his influential article from 2001 [150].

$$\gamma = 13/3$$

Of course, the above scaling can only hold in a finite frequency window as otherwise the energy carried in the GWB spectrum were unbounded: At an initial state of the inspiral, the hardening of the binary will be mostly driven by environmental effects, such that the spectrum will bend down towards lower spectral amplitudes at low frequencies. The high-frequency part of the spectrum is emitted by the inspiral of binaries just before coalescence. The highest possible GW frequency emitted through the inspiral of a binary hence corresponds to the so-called innermost stable circular orbit and is given by $f_{\text{ISCO}}^r = 4.7 \cdot 10^{-7} \text{ Hz} (10^{10} M_\odot/M)$, cf. eq. (16.78) in ref. [37]. For a monochromatic mass spectrum of SMBHBs, this would correspond to a sharp cutoff of the spectrum lying above the PTAs frequency band. In reality, however, the number density distribution dn/dz in eq. (3.31) will depend on the distribution of SMBHB masses, softening the decrease of the spectrum at high frequencies. Special attention needs to be paid due to the finiteness of the distribution of black holes with extremely high masses: there needs to be at least one source in order to emit a GW signal. We will investigate this issue in more detail in chapter 7, see e.g. fig. 7.2.

$$\begin{aligned} &\text{Deviations from} \\ &\gamma = 13/3 \end{aligned}$$

The NANOGrav collaboration uses the code **Holodeck** [7] to generate GWB spectra from SMBHBs based on realistic astrophysical distributions, starting from galaxy merger rates and black hole-host galaxy mass relationships. Recently, a phenomenological model for those GWB spectra has been put forward which allows the inference of black hole properties [149]. In doing so, it was found that several astrophysical quantities like the binary hardening timescale need to be beyond the borders of their previously expected ranges in order to explain the observed GWB spectrum at nHz frequencies [149]. In fig. 3.4, a comparison

$$\begin{aligned} &\text{NANOGrav vs.} \\ &\text{SMBHBs} \end{aligned}$$

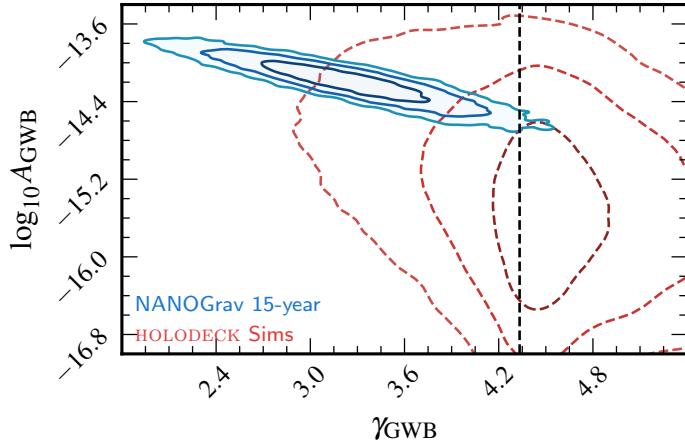


Figure 3.4: Marginalized posterior distributions of the amplitude A and spectral index γ for power-law fit of the HD-correlated red noise in the NANOGrav 15yr data set compared to predictions obtained using the `Holodeck` code for an SMBHB-induced background. The vertical line at $\gamma = 13/3$ refers to a population of SMBHBs whose inspiral is driven only by GW emission. This figure was taken from ref. [106].

of the expected amplitude and spectral tilt of the spectrum based on `Holodeck` simulations and the regions of parameter space preferred by the NANOGrav 15yr data set can be found. Even though the tension between the contours of prediction and expectation still depends on the used priors of the astrophysical simulations, the figure already indicates that there is room (if not the necessity) for an additional GW source which also contributes to the observed nHz signal. This serves as a source of motivation when searching for alternative explanations of the signal in the following chapters 5 and 7.

4

DARK SECTOR PHASE TRANSITIONS

*It cannot be seen, cannot be felt,
Cannot be heard, cannot be smelt,
It lies behind stars and under hills,
And empty holes it fills.*

— GOLLUM

In the previous two chapters we have introduced the field of GW cosmology and reviewed the state of PTA searches for stochastic GWBs. Now, we will go beyond the established Λ CDM expansion history and beyond SM physics by studying FOPTs in the early universe. As we have seen in the introductory section 2.1 on the chronology of the early cosmos, the history of our universe can be understood as a history of subsequent PTs: during inflation, a scalar field slowly rolled to obtain a new vacuum expectation value (vev). During reheating the equation of state of the Universe then changed from vacuum to radiation, corresponding to yet another PT. It followed the EWPT, as well as chiral symmetry breaking and confinement during the QCD PT. As we will show below, PTs are a generic feature of QFTs at finite temperature. This principle also applies to additional gauge groups extending the SM of particle physics: as noted in section 2.2, only 15% of the total non-relativistic matter in our Universe is accounted for by the SM particle content. The remaining 85% of matter motivate us to study the spontaneous breaking of a *dark* symmetry group in this chapter.

In the following section 4.1 we first make an analogy to a simple quantum mechanical harmonic oscillator coupled to a thermal bath to get an intuition for how a quantum field behaves at finite temperature. What follows is a discussion of the effective potential and the thermodynamics of FOPTs in the early universe in section 4.2. In section 4.3 we then come to the prediction of GWBs from PTs in DSs.

*Adding a dark PT
to the series of SM
ones*

*Summary of this
chapter*

4.1 FINITE-TEMPERATURE EFFECTS IN QFT

4.1.1 A quantum harmonic oscillator in a thermal bath

To start our discussion of thermal field theory, we will begin with the conceptually much easier case of a quantum mechanical harmonic oscillator with Hamiltonian \mathcal{H} , the trace of which defines the partition function $Z(T) \equiv \text{Tr} [e^{-\beta \mathcal{H}}]$. Here and in the following, $\beta = T^{-1}$ denotes

the inverse temperature of the bath to which the harmonic oscillator (and later the QFT) is coupled. For a frequency ω of the oscillator, we obtain the partition function [151]

$$Z_{\text{ho}} = \begin{cases} \sum_{n=0}^{\infty} \exp[-\beta\omega(n + \frac{1}{2})] = e^{-\beta\omega/2} (1 - e^{-\beta\omega})^{-1} \\ \sum_{n=0}^1 \exp[-\beta\omega(n + \frac{1}{2})] = e^{\beta\omega/2} (1 + e^{-\beta\omega/2}), \end{cases} \quad (4.1)$$

where the above (below) expression holds for a bosonic (fermionic) oscillator. One can show that a system in thermal equilibrium will eventually go to a state with minimal free energy $F \equiv -T \ln Z(T)$. For the above partition function of a harmonic oscillator, we obtain

$$F_{\text{ho}} = \begin{cases} \frac{\omega}{2} + T \ln(1 - e^{-\beta\omega}) & \text{for bosons} \\ -\frac{\omega}{2} - T \ln(1 + e^{-\beta\omega}) & \text{for fermions.} \end{cases} \quad (4.2)$$

*The free energy of
a harmonic
oscillator*

This is already the result which we wanted to obtain in this brief section: the free energy of a single harmonic oscillator is the sum of a temperature-independent term corresponding to the ground state energy $\hbar\omega/2$ and a temperature-dependent term which accounts for the occupation of higher states due to the available thermal energy. The first term is purely quantum mechanical and corresponds to Heisenberg's uncertainty in the oscillator's position and momentum, whereas the second term instead arises due to the coupled thermal bath and dominates the free energy for sufficiently high temperatures. Indeed, we will recover a remarkably similar expression after a much more tedious computation for the case of QFT coupled to a thermal bath.

4.1.2 The Kubo-Martin-Schwinger relation

We now want to transition from quantum mechanics to QFT. To do so, let us repeat that in statistical mechanics the expectation value of a given operator A is given by the thermally averaged sum of A over eigenstates of the Hamiltonian \mathcal{H} ,

$$\langle A \rangle_T \equiv \frac{1}{Z} \text{Tr} [e^{-\beta\mathcal{H}} A]. \quad (4.3)$$

In particular, this also holds for the correlators of quantum fields ϕ . To obtain the famous Kubo-Martin-Schwinger relation [152, 153], consider the two-point function of a quantum field $\phi_{\mathbf{y}}$ at position \mathbf{y} at time t and the field at position \mathbf{x} at zero time,

$$\begin{aligned} \langle \phi_{\mathbf{y}}(t) \phi_{\mathbf{x}}(0) \rangle_T &= \frac{1}{Z} \text{Tr} [e^{-\beta\mathcal{H}} \phi_{\mathbf{y}}(t) \phi_{\mathbf{x}}(0)] \\ &= \frac{1}{Z} \text{Tr} [\phi_{\mathbf{y}}(t) e^{-\beta\mathcal{H}} e^{i(-i\beta\mathcal{H})} \phi_{\mathbf{x}}(0) e^{-i(-i\beta\mathcal{H})}] \\ &= \frac{1}{Z} \text{Tr} [\phi_{\mathbf{y}}(t) e^{-\beta\mathcal{H}} \phi_{\mathbf{x}}(-i\beta)] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{Z} \text{Tr} \left[e^{-\beta \mathcal{H}} \phi_{\mathbf{x}}(-i\beta) \phi_{\mathbf{y}}(t) \right] \\
&= \langle \phi_{\mathbf{x}}(-i\beta) \phi_{\mathbf{y}}(t) \rangle_T \\
&= \pm \langle \phi_{\mathbf{y}}(t) \phi_{\mathbf{x}}(-i\beta) \rangle_T.
\end{aligned} \tag{4.4}$$

To arrive at the final expression we made repeated use of the permutation rule for traces and employed the time evolution $\phi_{\mathbf{x}}(t) = e^{i\mathcal{H}t} \phi_{\mathbf{x}}(0) e^{-i\mathcal{H}t}$. In the last step, a minus sign occurs only for fermionic fields due to them anti-commuting, but not for bosonic fields. We can hence see that a bosonic (fermionic) field ϕ needs to be symmetric (anti-symmetric) and cyclic in time with the periodicity $-i\beta$ if it is coupled to a thermal bath with inverse temperature β .

The emergence of periodicity in imaginary time

We now go to *purely* imaginary time $t \rightarrow \tau = -it$ by performing a Wick rotation. We find that in the so-called imaginary-time (or Matsubara) formalism, the Euclidean time parameter τ can be identified with the inverse temperature β [154]. In performing the Wick rotation, we gained access to the description of systems in thermal equilibrium, but lost control over dynamical processes happening in real time. In the alternative, more involved real-time (or Keldysh) formalism [155] both is possible simultaneously, thus allowing the description of out-of-equilibrium dynamics. For our goal of identifying the phase structure of a given QFT, the Matsubara formalism is sufficient, however. To go from correlators in a QFT at zero temperature to those of a finite-temperature QFT, we therefore need to replace the integral over the time component in the occurring integrals over Euclidean four-momentum space $k_E^\mu = (k_E^0, \mathbf{k}_E)$ with a discrete sum over the so-called Matsubara frequencies $k_E^0 = \omega_n$ [156],

$$\int \frac{d^4 k_E}{(2\pi)^4} f(k_E) \rightarrow T \sum_{n=-\infty}^{\infty} \int \frac{d^3 k}{(2\pi)^3} f(\omega_n, \mathbf{k})$$

with $\omega_n = \begin{cases} 2n\pi T & \text{for bosons} \\ (2n+1)\pi T & \text{for fermions.} \end{cases}$

$$\tag{4.5}$$

Matsubara sums

The summation corresponds to the integration over time in the theory with $T = 0$ and may be arbitrarily complicated for a general integrand. Intuitively, the Matsubara sum adds up all contributions to a given integral allowed by the field boundary conditions set by the Kubo-Martin-Schwinger relation in equation (4.4).

4.1.3 The effective potential of an abelian gauge theory

THE TREE-LEVEL POTENTIAL For discreteness, we now want to study the finite-temperature behavior of an additional dark $U(1)'$ extending the gauge groups of the SM. This model will be of particular relevance in chapter 6. Its Lagrangian density is given by

$$\mathcal{L} = |D_\mu \Phi|^2 - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \mu^2 |\Phi|^2 - \lambda |\Phi|^4$$

A dark $U(1)'$ gauge symmetry

$$+ \chi_L^\dagger i\cancel{D} \chi_L + \chi_R^\dagger i\cancel{D} \chi_R - y \Phi \chi_L^\dagger \chi_R - y \Phi^* \chi_R^\dagger \chi_L. \quad (4.6)$$

We ignore portal
couplings

Expanding around
the tree-level
minimum

The complex scalar field Φ and the chiral fermions χ_L and χ_R are charged under the $U(1)'$ gauge group. Their $U(1)'$ charges satisfy $Q_\Phi = Q_{\chi_L} - Q_{\chi_R}$, which we realize through the choice $Q_\Phi = 1$, $Q_{\chi_L} = +1/2$ and $Q_{\chi_R} = -1/2$ in order to preserve the gauge symmetry of the Yukawa interaction term and keep the model free of anomalies. This choice allows us to express the interactions of the chiral fermions as those of a single Dirac bi-spinor field in the following. The field strength tensor is defined in terms of the dark photon field A'_μ as $B_{\mu\nu} = \partial_\mu A'_\nu - \partial_\nu A'_\mu$ and the corresponding covariant derivative reads $D_\mu = \partial_\mu + igA'_\mu$. In principle, also a mixing term $\propto \Phi^2 H^2$ with the SM Higgs field H and a kinetic mixing term $\propto F_{\mu\nu} B^{\mu\nu}$ with the SM photon $F_{\mu\nu}$ would be allowed. In order to be able to ignore their effects on the effective potential for simplicity, we assume the respective portal couplings to be sufficiently small [157–159]. We will revisit their effect in chapters 5 and 6 when discussing the decay of DSs in order to satisfy cosmological constraints.

The above Lagrangian features the tree-level potential $V_{\text{tree}}(\Phi) = -\mu|\Phi|^2 + \lambda|\Phi|^4$, having a $U(1)'$ -symmetric minimum at $|\Phi|^2 = \mu^2/\lambda$. At a given point in space, the field must be in a single vacuum state, whereas the above condition is satisfied by an infinite set of fields related by the gauge transformation $\Phi \rightarrow e^{i\alpha}\Phi$ with $\alpha \in [0, 2\pi)$. For convenience we choose the vacuum to lie on the real field axis, thus breaking the $U(1)'$ spontaneously. Here, *spontaneously* refers to the field direction with zero complex phase being arbitrary among the line of potential minima. Upon this choice, we can expand the Lagrangian around $\Phi = (\phi + v_\phi + i\varphi)/\sqrt{2}$ with ϕ and φ being real fields and $v_\phi \equiv \mu/\sqrt{\lambda}$, eventually obtaining

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \partial_\mu \phi \partial^\mu \phi + \frac{1}{2} \partial_\mu \varphi \partial^\mu \varphi - \frac{1}{4} A'_{\mu\nu} A'^{\mu\nu} - \frac{1}{2} m_\phi^2 \phi^2 + \frac{1}{2} m_{A'}^2 A'^2 \\ & - g A'_\mu [\varphi \partial^\mu \phi - \phi \partial^\mu \varphi - v_\phi \partial^\mu \varphi] + \frac{g^2}{2} \phi^2 A'^2_\mu + \frac{g^2}{2} \varphi^2 A'^2_\mu \\ & + g^2 v_\phi \phi A'^2_\mu - \lambda v_\phi \phi^3 - \lambda v_\phi \varphi^2 \phi - \frac{\lambda}{4} \phi^2 \varphi^2 - \frac{\lambda}{4} \phi^4 - \frac{\lambda}{4} \varphi^4 \\ & + i\bar{\chi} \cancel{D} \chi - m_\chi \bar{\chi} \chi + \frac{g}{2} \bar{\chi} \cancel{A}' \gamma^5 \chi - \frac{y}{\sqrt{2}} \phi \bar{\chi} \chi + i \frac{y}{\sqrt{2}} \varphi \bar{\chi} \gamma^5 \chi \end{aligned} \quad (4.7)$$

with the tree-level masses

$$m_\phi^2 = 3\lambda v_\phi^2 - \mu^2 = 2\lambda v_\phi^2, \quad m_{A'}^2 = g^2 v_\phi^2, \quad (4.8a)$$

$$m_\varphi^2 = \lambda v_\phi^2 - \mu^2 = 0, \quad \text{and} \quad m_\chi^2 = \frac{y^2 v_\phi^2}{2}. \quad (4.8b)$$

INCLUDING QUANTUM EFFECTS The above mass spectrum only holds at tree-level, i.e. in the limit $\hbar \rightarrow 0$. Due to Φ being a quantum field, however, quantum corrections become relevant for non-zero field values. Mathematically, the effect of quantum corrections can be included by promoting the tree-level potential V_{tree} in eq. (4.6) to a so-called effective potential. The effective potential contribution of ℓ -loop Feynman

diagrams, corresponding to an expansion in powers of \hbar^ℓ , can be derived using the path-integral formulation of QFT and reads [160]

$$V_{\text{eff}}^\ell(\phi) = - \sum_{k=0}^{\infty} \frac{\phi^k}{k!} \Gamma_\ell^{(k)}(p=0). \quad (4.9)$$

Here, $\Gamma_\ell^{(k)}$ is the k -point, ℓ -loop effective vertex with all external momenta set to zero. Ignoring loop diagrams ($\ell = 0$), one therefore just recovers the tree-level potential. At leading loop order ($\ell = 1$) and only considering Φ -loops, we instead obtain the sum

$$\begin{aligned} V_{\text{eff},\Phi}^1(\phi) &= \left[\phi^2 \text{---} \text{---} + \phi^4 \text{---} \text{---} \text{---} \text{---} + \phi^6 \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} + \dots \right]_{p=0} \\ &= 2i \sum_{n=1}^{\infty} \int \frac{d^4 k}{(2\pi)^4} \frac{1}{2n} \left[\frac{6\lambda\phi^2/2}{k^2 + \mu^2 + i\epsilon} \right]^n. \end{aligned} \quad (4.10)$$

In the above diagrams we denote the real part ϕ of the Higgs field with dashed lines, whereas solid lines refer to Φ comprising both the imaginary and real field components. In the second step we used that the n -th diagram has $2n$ external legs and n propagators. Additionally, there are symmetry factors of $1/(2n)$ and $1/2^n$ from cyclic and anti-cyclic permutations of the vertices as well as permutations of the external lines of each vertex, respectively. The initial factor 2 accounts for the two internal dofs of Φ on the loop. Further, we used that each vertex comes with a factor $-6i\lambda$ and each internal propagator brings a factor $i/(k^2 + \mu^2 + i\epsilon)$.

We can perform the sum over n by comparing with the series expansion of $\ln(1+x)$ around $x = 0$. Further performing a Wick rotation we arrive at

$$V_{\text{eff},\Phi}^1(\phi) = \int \frac{d^4 k_E}{(2\pi)^4} \ln [k_E^2 + m_\phi^2(\phi)]. \quad (4.11)$$

So far we only considered the effect of quantum corrections arising due to the self-coupling λ of Φ to ϕ , but not the effect of the gauge boson A and the fermion χ . One can show that also fermions and gauge bosons contribute in a similar way to the effective potential through their loop contributions, up to a factor corresponding to their dofs and a respective sign $\eta_a = +1$ (-1) for bosons (fermions) [156]:

$$V_{\text{eff}}^1(\phi) = \sum_a \frac{\eta_a n_a}{2} \int \frac{d^4 k_E}{(2\pi)^4} \ln [k_E^2 + m_a^2(\phi)]. \quad (4.12)$$

Here, the sum goes over all particle species a coupled to ϕ , with n_a being their internal dofs.

The one-loop order

Contributions from all charged fields

GOING TO FINITE TEMPERATURE To now compute the effective potential at finite temperature $T_d = \xi T$ (with the index d indicating a

potential mismatch ξ between the DS temperature and the visible sector temperature T), we make use of the replacement rule from eq. (4.5) and obtain

$$\begin{aligned} V_{\text{eff}}^1(\phi, T_d) &= \sum_a \frac{\eta_a n_a T_d}{2} \sum_n \int \frac{d^3 k}{(2\pi)^3} \ln [\omega_n^2 + k^2 + m_a^2] \\ &= \sum_a \eta_a n_a \int \frac{d^3 k}{(2\pi)^3} \left[\frac{E_a}{2} + T_d \ln \left(1 - \eta_a e^{-E_a/T_d} \right) \right] \\ &= V_{\text{CW}}(\phi) + V_T(\phi, T_d). \end{aligned} \quad (4.13)$$

To simplify the notation, the energy parameter $E_a^2 \equiv k^2 + m_a^2$ was introduced. What we retrieve, closely resembles the free energy of a single harmonic oscillator which we derived in eq. (4.2). Now, however, the frequency ω of the oscillator corresponds to a quantum field's energy E_a , which is integrated over. This is analogous to the naive image of a quantum field to be a superposition of infinitely many harmonic oscillators at each point in space. We obtained both a temperature-dependent and temperature-independent part, the latter of which corresponds to the (yet unrenormalized) Coleman-Weinberg potential [161].

The harmonic oscillator is everywhere

The temperature-dependent part can be expressed as

$$V_T(\phi, T_d) = \frac{T_d^4}{2\pi^2} \sum_a \eta_a n_a J_{\eta_a} \left(\frac{m_a^2(\phi)}{T_d^2} \right) \quad \text{with} \quad (4.14)$$

$$J_{\eta_a}(z^2) \equiv \int_0^\infty dy y^2 \ln \left[1 - \eta_a \exp \left(-\sqrt{y^2 + z^2} \right) \right]. \quad (4.15)$$

Fig. 4.1 shows the real part of the thermal functions J in dependence of $z_a = m_a^2/T_d^2$.¹ In the low-temperature limit, $z \rightarrow \infty$, the thermal functions and therefore also V_T vanish. In the high-temperature limit, $z \rightarrow 0$, the thermal functions can instead be approximated by [156]

$$J_{\text{bos}}(z^2) \approx -\frac{\pi^4}{45} + \frac{\pi^2}{12} z^2 - \frac{\pi}{6} z^3 - \frac{z^4}{32} \ln(z^2) + \text{const}, \quad (4.16a)$$

$$J_{\text{ferm}}(z^2) \approx \frac{7\pi^4}{360} - \frac{\pi^2}{24} z^2 - \frac{z^4}{32} \ln(z^2) + \text{const}. \quad (4.16b)$$

Interpreting the thermal corrections

The leading-order correction to the tree-level potential due to a single bosonic dof hence reads $\pi^2 T_d^4 / 90$ if $T_d \gg m_a$, i.e. the Stefan-Boltzmann law we already identified in eq. (2.14b). Typically, this contribution to the vacuum energy density is absorbed into the cosmological constant, as it does not directly influence the dynamics of the scalar field. Note, however, in particular the z^2 terms in eqs. (4.16), which translate to a term $V_T(\phi, T_d) \propto \phi^2 T_d^2$, acting as a temperature-dependent mass in the effective potential. At sufficiently high temperatures, these

1 The behavior of the thermal functions in fig. 4.1 at negative z^2 is relevant as the scalar field will obtain an imaginary mass wherever the potential has a negative curvature. An in-depth study of the imaginary parts of the effective potential was performed in ref. [162]. Generally speaking, an imaginary part of the effective potential indicates an instability of the vacuum at a specific field value, eventually triggering a PT.

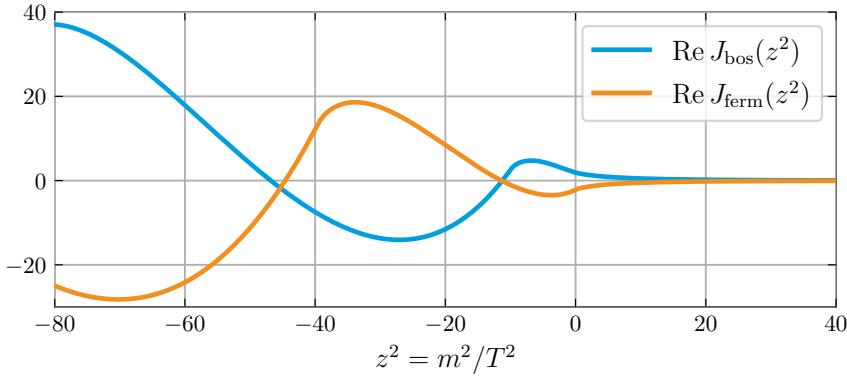


Figure 4.1: The real part of the thermal functions $J_{\text{bos}}(z^2)$ and $J_{\text{ferm}}(z^2)$.

temperature-dependent mass terms inevitably dominate over any other mass contribution, forcing the potential minimum to reside at vanishing field values $\phi = 0$. This phenomenon is known as symmetry restoration at high temperatures, if $\phi \neq 0$ refers to a phase in which a symmetry of the high-temperature QFT is broken as in the present case of the spontaneous $U(1)'$ breaking.

High-temperature symmetry restoration is a feature of many QFTs and the reason for us to understand the cosmic expansion history as a sequence of PTs: When temperature drops, the effect of V_T decreases and the positions of the potential minima change. In fact it is not of further relevance if Φ is a fundamental scalar (as the SM Higgs field triggering the EWPT) or a bi-linear condensate $\langle \bar{q}q \rangle$ of fermions (like the quarks q in the QCD PT). In the specific case of the tree-level potential $V_{\text{tree}} = -\mu^2\phi^2 + \lambda\phi^4$, again considering only the ϕ^2 term of the effective potential and ignoring $\mathcal{O}(1)$ factors, we see that at high temperatures it will effectively include a term that reads $V(\phi, T_d) \supset (-\mu^2 + \lambda T_d^2)\phi^2$, indicating that for $T_d \lesssim \mu/\sqrt{\lambda}$ the $\phi = 0$ phase will become unstable.

High-temperature symmetry restoration

THE BREAKDOWN OF PERTURBATIVITY The contributions V_{CW} and V_T to the effective potential can hence be regarded as corrections to the potential energy density of a scalar field ϕ . However, both terms are still plagued by divergences. This is expected as eq. (4.10) includes diagrams with two and four external legs for which the integrand is unbounded in the ultraviolet (UV) limit of large loop momenta. But also the second, temperature-dependent term is expected to feature a breakdown of perturbativity by the mere existence of two energy scales, namely temperature and the scalar's mass [163]. The latter is an IR divergence which can be understood by the thermal distribution of bosons $[\exp(E/T) - 1]^{-1} \rightarrow T/E$ diverging for $T \gg E$: At high temperatures their distribution is dominated by low-energy bosons and diverges for soft bosons with $E \rightarrow 0$. This translates to an effective increase of the coupling between infrared bosons up to the point where the coupling

UV and IR divergences

between them becomes non-perturbative and our perturbative expansion breaks down.

The UV divergence of the Coleman-Weinberg potential can be treated by dimensional regularization. Using the $\overline{\text{MS}}$ renormalization scheme we arrive at [161, 164, 165]

$$V_{\text{CW}}(\phi) = \sum_x \eta_x n_x \frac{m_x^4(\phi)}{64\pi^2} \left[\ln \frac{m_x^2(\phi)}{\Lambda^2} - C_x \right] \quad (4.17)$$

with C_x being equal to $3/2$ ($5/6$) for scalars and fermions (gauge bosons) and Λ being the renormalization scale [156, 164, 165]. To arrive at the above expression, the dimensionality of the integral in eq. (4.13) was first formally shifted to $4 - \epsilon$. The singularities are then isolated in terms $\propto \epsilon^{-1}$, which need to be canceled by adding a finite counterterm potential V_{ct} [156]. In this thesis, we will impose the renormalization conditions $0 = \partial_\phi (V_{\text{CW}} + V_{\text{ct}})|_{\phi=\Lambda}$ and $0 = \partial_\phi^2 (V_{\text{CW}} + V_{\text{ct}})|_{\phi=\Lambda}$, effectively fixing the tree-level vev and the Higgs mass m_ϕ to remain unchanged when going from tree-level to one-loop order, and set $\Lambda = v_\phi$.

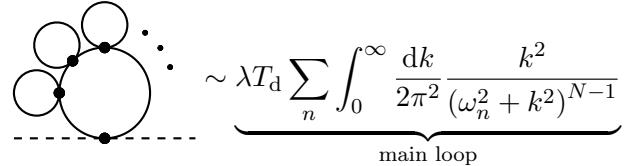
HARD THERMAL LOOPS The divergences of the T_d -dependent term in eq. (4.13) cannot be treated as simply: Consider for instance diagram (a) in fig. 4.2. Evaluating it by applying the replacement rule from eq. (4.5) for vanishing external momenta we obtain

$$\begin{aligned} \Pi_\phi(T_d) &= \lambda T_d \sum_n \int \frac{d^3 k}{(2\pi)^3} \frac{1}{\omega_n^2 + k^2 + m_\phi^2} \\ &= \lambda \int \frac{d^4 k}{(2\pi)^4} \left(\frac{1}{2E_\phi} + \frac{1}{E_\phi} \frac{1}{e^{-E_\phi/T_d} - 1} \right) \stackrel{T \gg m_\phi}{\approx} \frac{\lambda T_d^2}{12} \end{aligned} \quad (4.18)$$

Debye masses

In the last step we assumed that $T_d \gg m_\phi$ and discarded the temperature-independent and UV-divergent first term in brackets. In fig. 4.2 the other one-loop corrections for the scalar and gauge boson propagator of the Lagrangian in eq. (4.6) are listed. These so-called Debye masses are generally of the form $\Pi_\phi \propto \lambda T_d^2$ for scalars and $\Pi_{A'} \propto g^2 T_d^2$ for longitudinally polarized gauge bosons. We left out fermion Debye masses, since they will not appear in the effective potential at leading order, as we will show below.

DAISY DIAGRAMS We will now attach $N - 1$ more Φ loops to the central Φ loop of the diagram (a) in fig. 4.2 to obtain the daisy diagram, which contributes to the mass of the real field ϕ as [166]



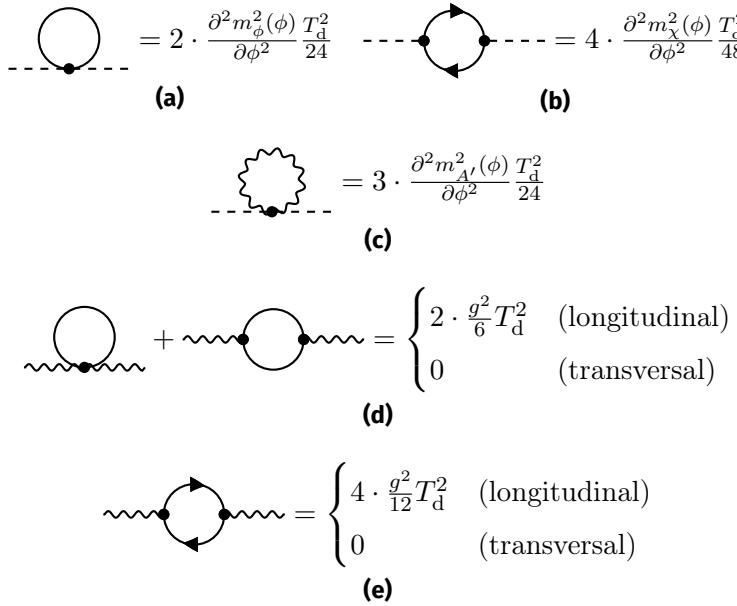


Figure 4.2: The so-called hard thermal loop corrections to the scalar (a, b and c) and gauge boson propagators (d and e) of the $U(1)'$ model in eq. (4.6) together with their Debye masses in the limit $T_d \gg m_a$, where m_a is the mass of the particle constituting the loop.

$$\times \underbrace{\left[\lambda T_d \sum_n \int_0^\infty \frac{dk}{2\pi^2} \frac{k^2}{\omega_n^2 + k^2} \right]^{N-1}}_{\text{petals}} \sim \frac{(\lambda T_d) (\lambda T_d^2)^{N-1}}{\mu^{2N-3}} = \alpha^{N-3/2} \lambda^{3/2} T_d^2. \quad (4.19)$$

In the last step, we introduced the effective coupling $\alpha \equiv \lambda T_d^2 / \mu^2$. We can immediately see that the main loop is IR divergent for $N > 2$ due to the Matsubara zero-mode $\omega_n = 0$. This phenomenon can only happen for bosonic thermal loops, which is the reason why we do not need to consider fermionic loops at leading order. At the temperature $T_d \sim \mu/\sqrt{\lambda}$, which we motivated above by the mass μ^2 and Π_ϕ becoming of the same order, eventually triggering a PT, the factor $\alpha \sim 1$. This implies that $\Pi_{\text{daisy}} \propto \lambda^{3/2} T_d^2$. Consequently, adding another petal to the daisy does not alter its mass contribution, which strongly contradicts our naive expectation from zero-temperature QFT, where in the case of perturbative couplings λ , a higher number of loops in a Feynman diagram typically corresponds to a smaller contribution. Instead, in thermal field theory, the daisy diagram contributes with a term proportional to $\lambda^{3/2}$ to the effective potential, thereby constituting the next-to-leading order correction to the effective temperature-dependent potential.

Picking daisies is important

RESUMMING HARD THERMAL LOOPS The daisy diagram showed to be important in particular around the PT scale $T_d \simeq \mu/\sqrt{\lambda}$ and to

diverge in the IR for $N > 2$ total loops. To now account for its effect in the effective potential, the propagators for bosons with tree-level mass m_ϕ have to be resummed. This can be achieved by adding up propagators with an increasing number of hard thermal loops with Debye mass $\Pi_\phi(T_d)$,

$$\frac{1}{p^2 - m_\phi^2} + \frac{\Pi_\phi}{(p^2 - m_\phi^2)^2} + \frac{\Pi_\phi^2}{(p^2 - m_\phi^2)^3} + \dots = \frac{1}{p^2 - m_\phi^2 - \Pi_\phi}$$

$$\text{---} + \text{---} \circ \text{---} + \text{---} \circ \text{---} + \dots = \text{---} \bullet \text{---} . \quad (4.20)$$

Effectively, the resummation hence boils down to the substitution $m_\phi^2 \rightarrow m_\phi^2 + \Pi_\phi(T_d)$ in $V_T(\phi, T_d)$. In this thesis, we choose to use the Arnold-Espinoza method [167] for daisy resummation in which only the Matsubara-zero mode's contribution is resummed, as opposed to the Parwani method [168] or the full dressing procedure [163], to spare the introduction of temperature-dependent counter terms. Algebraically, the resummation of the hard thermal loops can be achieved by the replacement [169]

$$[V_{\text{CW}} + V_T]_{m_\phi^2 \rightarrow m_\phi^2 + \Pi_\phi} = V_{\text{CW}} + V_T + V_{\text{daisy}}$$

$$\text{with } V_{\text{daisy}}(\phi, T_d) = -\frac{T_d}{12\pi} \left[(m_\phi^2 + \Pi_\phi(T_d))^{3/2} - (m_\phi^2)^{3/2} \right] \quad (4.21)$$

The contributions from gauge bosons follow the same expression up to a factor for their internal dofs. As only longitudinal gauge bosons obtain Debye masses (see fig. 4.2), only the longitudinal dofs of a gauge boson need to be considered.

SUMMARY Summing all discussed terms together, we obtain the one-loop, daisy-resummed effective potential of the QFT defined by the Lagrangian in eq. (4.6) [170]

$$V_{\text{eff}}(\phi, T_d) = V_{\text{tree}} + V_{\text{CW}} + V_{\text{ct}} + V_T + V_{\text{daisy}} \quad (4.22)$$

with the individual contributions

$$V_{\text{CW}}(\phi) = \sum_{a=\phi, \varphi, A', \chi} \eta_a n_a \frac{m_a^4(\phi)}{64\pi^2} \left[\ln \frac{m_x^2(\phi)}{v_\phi^2} - C_a \right],$$

$$V_T(\phi, T_d) = \frac{T^4}{2\pi^2} \sum_{a=\phi, \varphi, A', \chi} \eta_a n_a J_{\eta_a} \left(\frac{m_a^2(\phi)}{T_d^2} \right),$$

$$V_{\text{daisy}}(\phi, T_d) = -\frac{T_d}{12\pi} \sum_{b=\phi, \varphi, A'_L} n_b \left[(m_b^2 + \Pi_b(T_d))^{3/2} - (m_b^2)^{3/2} \right],$$

$$V_{\text{ct}}(\phi) = -\frac{\delta\mu^2}{2} \phi^2 + \frac{\delta\lambda}{4} \phi^4$$

$$\begin{aligned}
\text{with } \Pi_\phi = \Pi_\varphi &= \left(\frac{\lambda}{3} + \frac{y^2}{12} + \frac{g^2}{4} \right) T_d^2, \quad \Pi_{A'} = \frac{3}{4} g^2 T_d^2, \\
\delta\mu^2 &= \left[\frac{3}{2\phi} \frac{dV_{CW}(\phi)}{d\phi} - \frac{1}{2} \frac{d^2V_{CW}(\phi)}{d\phi^2} \right] \Big|_{\phi=v_\phi}, \\
\text{and } \delta\lambda &= \left[\frac{1}{2\phi^3} \frac{dV_{CW}(\phi)}{d\phi} - \frac{1}{2\phi^2} \frac{d^2V_{CW}(\phi)}{d\phi^2} \right] \Big|_{\phi=v_\phi}.
\end{aligned} \tag{4.23}$$

As above, n_a are the dofs of the fields coupled to ϕ , η_x is $+1$ (-1) for bosons (fermions), $C_a = 3/2$ ($5/6$) are the renormalization constants for scalars and fermions (gauge bosons), and J_{η_a} are the thermal functions as defined in eq. (4.15).

There have been developments in favor of a more rigorous calculation of the effective potential which allows for an expansion not in terms of the number of loops but in terms of powers of the coupling constants. This can be achieved through the construction of a three-dimensional effective field theory by formally integrating out the hard thermal scale (corresponding to the resummation of daisies), as well as the parameterically smaller soft and ultra-soft thermal scale (also resumming so-called super-daisies, lollipops and sunset diagrams) [163, 166]. The resulting effective potentials recovered through this procedure in particular do not suffer from a strong dependence of a specific chosen gauge and renormalization scale in contrast to the above calculation. Recently, dimensional reduction has become a feasible task using the Mathematica package **DRalgo** [171]. In chapter 6 we use **DRalgo** in order to perform a cross-check and validate our results obtained using the effective potential from eq. (4.22).

*Super-daisies,
lollipops and
sunsets*

4.2 BUBBLE NUCLEATION AND PERCOLATION

We are now well-equipped to study PTs in the early universe. To start with, let us study the derived effective potential in more detail: In fig. 4.3 we schematically plot $V_{\text{eff}}(\phi, T_d)$ for a range of temperatures given two benchmark points, one with $g = 0$ on the left-hand side and one with $g \neq 0$ on the right-hand side. At low temperature, in both cases the tree-level potential is recovered, having a minimum at the vev v_ϕ . At sufficiently high temperatures, we further see that in both cases the minimum of the potential shifts to $\phi = 0$, which corresponds to the previously mentioned phenomenon of symmetry restoration. For intermediate temperatures, the effective potentials however are fundamentally different: For non-zero gauge couplings a thermally induced potential barrier is produced.

*A thermally
induced barrier in
 V_{eff}*

The origin of this potential barrier can be traced down to the z^3 term in eq. (4.16a), which only appears in the bosonic but not the fermionic thermal function. In fact, one can show that a thermally induced potential barrier can only be produced in the presence of transversal gauge

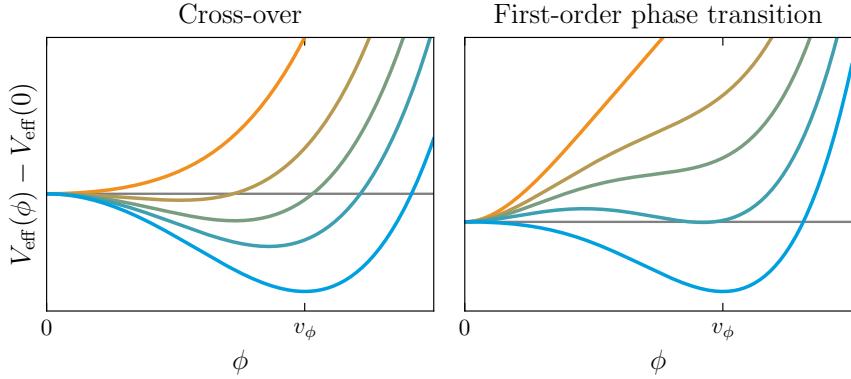


Figure 4.3: Sketch of the effective potential (4.22) corresponding to the Lagrangian in eq. (4.6) for a set of decreasing temperatures (orange to blue curves). Left: $g = 0$, no potential barrier can form. Right: $g \neq 0$, a FOPT can happen due to the generation of a potential barrier.

The barrier is due to transversal gauge bosons

Below $T_{d,c}$ bubbles can nucleate

The bubble nucleation rate

bosons, as the daisy potential V_{daisy} effectively cancels thermal barriers induced by loops of scalars and longitudinal gauge bosons: Expanding V_{daisy} in the high-temperature regime, it evaluates to $T_d^4 \sum_b z_b^3 / (12\pi)$, such that after adding the z_b^3 piece of V_T to it only the transversal gauge boson barrier survives.

In the presence of a potential barrier, the scalar field ϕ cannot smoothly transition to a new potential minimum, i.e. ϕ cannot be homogeneous at all times. Instead, for temperatures below the critical temperature $T_{d,c}$ at which the minima degenerate, bubbles of the new phase can nucleate, expand, accelerate up to relativistic velocities, and eventually collide, thereby possibly emitting strong gravitational radiation. In fig. 4.4 a sketch of the bubble nucleation can be found: A bubble's radius r , initially sufficiently large such that the bubble does not collapse under its own surface tension, will grow like $r = |\mathbf{x}| = \sqrt{R^2 + c^2 t^2}$ if the interactions of the bubble wall with the plasma can be neglected [172, 173].

We now want to compute the temperature of the primordial plasma at which bubble nucleation happens. There are two different mechanisms through which the transition can be triggered: quantum tunneling and thermal fluctuations. We will only consider the case of PTs induced by thermal fluctuations, which typically have a bubble nucleation rate which exceeds that of tunneling-induced transitions except for the case of very strong supercooling [165]. In a semi-analytical approximation the bubble nucleation rate (per unit volume and time) reads [172–174]

$$\Gamma(t) = A(T_d) \exp \left[-\frac{S_3(T_d)}{T_d} \right], \quad (4.24)$$

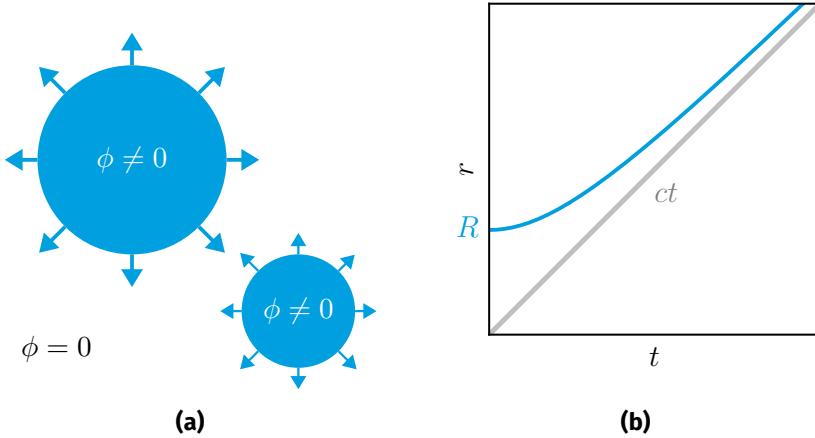


Figure 4.4: Left: Sketch of the bubble nucleation. Inside the bubbles $\phi \neq 0$, whereas the vacuum around them is still in the $\phi = 0$ phase. Right: Sufficiently large bubbles can expand and accelerate up the speed of light.

where $S_3(T_d)$ is referred to as the (Euclidean) bounce action. The latter is given by

$$S_3(T_d) \equiv S_3[\phi_b(\mathbf{x}; T_d)] = \int d^3x \left[\frac{(\nabla\phi_b)^2}{2} + V_{\text{eff}}(\phi_b, T_d) \right] \quad (4.25)$$

where $\phi_b(\mathbf{x}; T_d)$ is the “bounce” solution of the $O(3)$ -symmetric Klein-Gordon equation

$$\frac{\partial^2 \phi}{\partial r^2} + \frac{2}{r} \frac{\partial \phi}{\partial r} = \frac{dV_{\text{eff}}(\phi, T_d)}{d\phi} \equiv V'_{\text{eff}}(\phi, T_d). \quad (4.26)$$

Here, we made the bubble’s $O(3)$ symmetry manifest by choosing spherical coordinates with $r = |\mathbf{x}|$ denoting the distance to the bubble center. The latter Klein-Gordon equation interpolates $\phi(r)$ between the two phases and hence determines the bubble profile $\phi_b(r; T_d)$. To determine it numerically given a potential V_{eff} , we impose the boundary conditions $\phi(r \rightarrow \infty) \rightarrow 0$ and $\phi'(r = 0) = 0$ and use a shooting algorithm (see fig. 11 in ref. [165]).

The time-dependence in eq. (4.24) was reintroduced to indicate that $T_d(t)$ is a dynamical quantity due to the Hubble expansion. Note that previously we assumed that there are no dynamical processes such that we could work in the imaginary time formalism to derive $V_{\text{eff}}(\phi, T_d)$. This assumption still holds very well in the limit of a sufficiently slow Hubble expansion. This assumption can be checked explicitly by comparing typical particle interaction rates with the Hubble rate (cf. chapter 6.4.2).

*The reappearance
of dynamics*

In fig. 4.5 a schematic overview of the different quantities going into the calculation of the bubble nucleation rate is shown: From a given effective potential the phase structure can be obtained by tracking the position of the potential minima over a range of temperatures. If there exists an

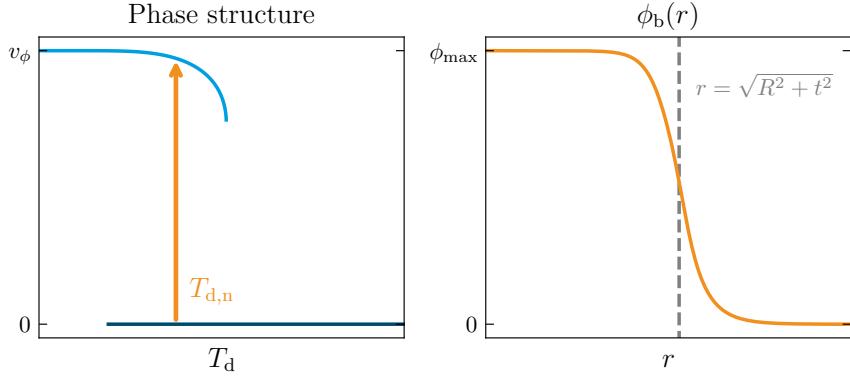


Figure 4.5: *Left:* The coexistence of potential minima in range of temperatures allows for a FOPT which succeeds through the nucleation of bubbles at T_n . *Right:* A bubble profile which interpolates between the vacuum states on both sides of the bubble wall at $\sqrt{R^2 + t^2}$.

The bubble profile

overlap in which two phases separated by a potential barrier coexist, a FOPT can occur. The bubble profile of an expanding bubble is shown on the right: Within the bubble, at $r < \sqrt{R^2 + t^2}$, where R is the initial bubble radius and t is the time passed since the individual bubble's nucleation, the scalar field obtains the vev of the true minimum. In front of the bubble wall, the field quickly goes to the false vacuum's vev. Note that there is some temperature dependence of the true minimum's vev as can be seen in the light blue curve in the left panel, which we refer to ϕ_{\max} in the right panel.

We define the point in time at which nucleation happens by requiring that one bubble per Hubble patch has been produced. By dimensional analysis and for simplicity the prefactor of the nucleation rate can be approximated as $A(T_d) \sim T_d^4$. This is usually sufficient for leading-order calculations due to the exponential dependence of the bounce action. For a more rigorous and already automated calculation of $A(T_d)$ up to one-loop order see for instance ref. [175]. The number of bubbles per Hubble patch can be calculated by integrating $\Gamma(t)$ over time. Due to the exponential time dependence, this integral can be well approximated by only considering the lowest temperature at which Γ is evaluated, i.e. the nucleation temperature. In doing so we immediately obtain the nucleation condition $\Gamma H^{-4}|_{t_n} = 1$, which can be rewritten as [27]

$$\frac{S_3(T_d)}{T_d} \Big|_{T_{d,n}=\xi_n T_n} \simeq 146 - 2 \ln \left(\frac{g_*(T_n)}{100} \right) - 4 \ln \left(\frac{T_n}{100 \text{ GeV}} \right) \quad (4.27)$$

Nucleation happens when $S_3(T_d)/T_d \simeq \mathcal{O}(150)$

upon plugging in the Hubble rate $H^2(T) = \frac{\pi^2}{90} g_*(T) T^4 / m_{\text{Pl}}^2$ in radiation domination. In summary, nucleation is expected to happen when S_3/T_d crosses $\log(m_{\text{Pl}}^4/T_n^4) \simeq \mathcal{O}(150)$. Numerically, the nucleation temperature can hence be found iteratively by repeated evaluation of the above nucleation condition at different temperatures. Note that the above

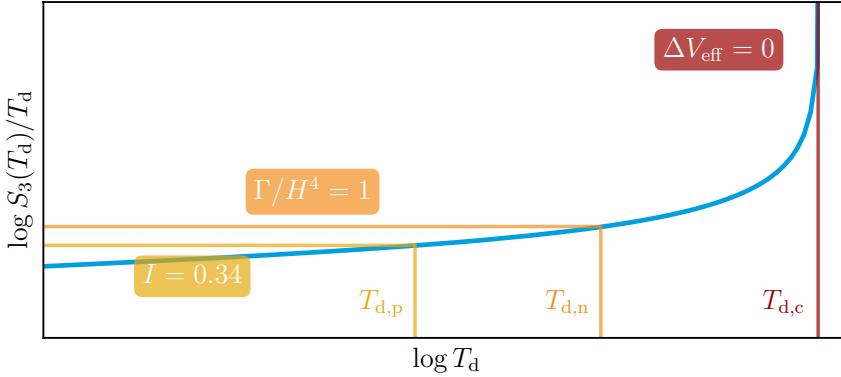


Figure 4.6: The temperature dependence of the bounce action $S_3(T_d)/T_d$ over a range of temperatures. At $T_{d,c}$ the bounce action diverges since $\Delta V_{\text{eff}} = 0$, corresponding to a vanishing bubble nucleation rate. At $T_{d,n}$ the bounce action becomes low enough for the nucleation of one bubble per Hubble volume. Percolation happens at $T_{d,p}$.

criterion is only approximate and needs to be corrected in case the amount of vacuum energy in the DS is so large that it leads to a second inflationary period, breaking the assumption of radiation domination.

In fig. 4.6 we schematically show how the bounce action drops for a decreasing DS temperature. Note that at $T_{d,c}$ when the two minima are at the same vacuum energy such that $\Delta V_{\text{eff}} = 0$, the bounce action diverges and the bubble nucleation rate hence drops to zero, corresponding to a zero probability for the nucleation of a bubble at a given point in space.

Bubble nucleation alone is not sufficient for the completion of a PT. In case of a strongly super-cooled PT, i.e. for a large difference in vacuum energy between the two phases, the Hubble parameter can be dominated by ΔV_{eff} leading to an exponential growth of the scale factor, forbidding the collision of bubbles. It therefore makes sense to check whether at some point in time the new phase actually percolates the universe by a connected web of bubbles [176–178]. This point in time is specified by roughly 30% of the Hubble volume having transitioned to the true vacuum state. The corresponding false-vacuum fraction is given by $P(T) = \exp[-I(T)]$ with [165]

$$I(T) = \frac{4\pi}{3} v_w^3 \int_T^{T_c} dT' \frac{\Gamma(T')}{T'^4 H(T')} \left(\int_T^{T'} \frac{dT''}{H(T'')} \right)^3, \quad (4.28)$$

where v_w is the terminal bubble wall velocity resulting from interactions of the bubble wall with the surrounding plasma. The resulting percolation temperature $T_p \lesssim T_n$, which can be determined from the condition $I(T_p) = 0.34$, approaches the nucleation temperature T_n in the limit of weak PTs, but can be orders of magnitude smaller in the case of strong super-cooling [179].

Percolation

4.3 GRAVITATIONAL WAVES FROM DARK SECTOR PHASE TRANSITIONS

In principle, we now have all the necessary ingredients to predict the GWB emitted during a FOPT. In practice, this is typically achieved by simulating the nucleation, expansion, and collision of bubbles on a lattice. This approach allows us to access the anisotropic stress within the perturbed plasma, and thereby the GW spectrum that emerges over time [151]. However, bubble expansion is complicated by the feedback from fluid dynamics, making these computations increasingly demanding depending on the size of the simulated volume and the chosen grid spacing. To accurately approximate the generated GW signal, the simulated volume should include as many bubbles as possible, and the grid spacing must be smaller than the relevant scales to precisely model the fluid dynamics. Additionally, the bulk fluid motion that sources GWs can persist much longer than the bubble collisions themselves [178], necessitating long simulation times. As a result, predicting the GW spectrum for a specific transition in a given effective potential is computationally too expensive to simulate in each case. Fortunately, there are widely used approximations derived from a combination of analytical and numerical methods that can effectively predict GWB spectra. We will now discuss how to map a given effective potential to these semi-analytical expressions.

To do so, let us introduce a set of useful dimensionless parameters which characterize the thermodynamics of dark sector phase transitions (DSPTs): The most relevant parameters for the calculation of the GWB spectrum, next to the previously derived percolation temperature T_p , are the strength parameters α and α_{DS} and the speed β/H of the PT.

The strength parameter α quantifies the amount of latent heat released in the PT and relates it to the total energy density $\rho_{SM}^P + \rho_{DS}^P$ of both the DS and the SM bath at the time of percolation. In ref. [180], the trace $\theta_d \equiv -\eta_{\mu\nu} T_d^{\mu\nu} = -4V_{eff} + T_d \frac{\partial V_{eff}}{\partial T_d}$ of the DS's energy momentum tensor $T_d^{\mu\nu}$ has been advertised to be a good parameterization for mapping between a given effective potential to the performed numerical simulations. The effective potential is here understood to be evaluated at the percolation temperature $T_{d,p}$. The latent heat can then be obtained through $\Delta\theta_d/4$, where Δ refers to a difference between the two phases. The strength of the phase transition can hence be expressed as

$$\alpha \equiv \frac{\Delta\theta_d/4}{\rho_{SM}^P + \rho_{DS}^P} \quad \text{with} \quad \Delta\theta_d = \left[-4\Delta V_{eff} + T_d \frac{\partial \Delta V_{eff}}{\partial T_d} \right]_{T_{d,p}} > 0, \quad (4.29)$$

where ΔV_{eff} denotes the potential difference between the two vacua. This parameterization of α has become a useful standard as it does not suffer from the inaccuracies which arise when using the pressure or vacuum energy difference instead of the trace difference $\Delta\theta_d$, as advertised in earlier works [181, 182].

*Lattice simulations
are expensive*

*The transition
strength*

In case the DS is completely secluded, there is another important transition strength parameter: It can be obtained by instead normalizing $\Delta\theta_d$ to the DS energy density alone [1, 4],

$$\alpha_{DS} \equiv \frac{\Delta\theta_d/4}{\rho_{DS}^p} = \alpha \left(1 + \frac{1}{\xi_p^4} \frac{g_{SM}(T_p)}{g_{DS}(T_p)} \right), \quad (4.30)$$

where $\xi_p = T_{d,p}/T_p$ is the temperature ratio between symmetric phase of the DS and the SM bath at the time of percolation. In a DSPT, the bubble wall can only feel the effect of DS states, as visible sector states do not obtain masses throughout the transition. The percentage of liberated vacuum energy density available for accelerating the bubble wall will hence depend on α_{DS} and not on α . This is an important difference with respect to the case of visible sector PTs. Also quantitatively, the difference between α_{DS} and α can be enormous due to the factor ξ_p^4 . The temperature ratio ξ_p therefore strongly impacts the dynamics of the dark bubble walls.

The speed of the PT expressed in units of the Hubble rate at the time of percolation can be determined by calculating the slope of the bounce action at T_p ,

$$\beta/H \equiv \frac{1}{H} \frac{d \log \Gamma}{dt} \Big|_{t_p} \simeq T_{d,p} \frac{d}{dT_d} \frac{S_3(T_d)}{T_d} \Big|_{T_{d,p}}. \quad (4.31)$$

Typically, S_3/T_d can be expanded as a polynomial in T_d around $T_{d,p}$, such that $\beta/H = \mathcal{O}(S_3(T_{d,p})/T_{d,p})$. As we argued in eq. (4.27), nucleation happens when $S_3(T_{d,n})/T_{d,n} = \mathcal{O}(150)$. For sufficiently weak phase transitions $T_{d,p} \simeq T_{d,n}$ holds. Quite generally, one can hence expect $\beta/H \simeq \mathcal{O}(100-1000)$ depending on the strength of the PT [1]. Obtaining smaller values for β/H , in order to produce stronger GW signals, is possible for near-conformal potentials [183–190] or at the expense of tuning the scalar potential such that the system is close to meta-stable (if it does not tunnel during the lifetime of the Universe). For as small values as $\beta/H = \mathcal{O}(1)$, however, the phase transition might not complete and a phase of eternal inflation might occur [165, 183, 184, 187–190].

Typically, a PT is referred to as *strong* when $\alpha \gtrsim \mathcal{O}(1)$ and *slow* when $\beta/H \lesssim \mathcal{O}(100)$. Both conditions need to be fulfilled in order to obtain observable GW signals as we will see in the following section. This corresponds to a latent heat release $\Delta V_{eff} \simeq \rho_{SM}^p + \rho_{DS}^p$ of the order of the energy density of the primordial plasma before the transition and a duration of the transition β^{-1} of at least a hundredth of a Hubble time. To give some meaning to these scales, let us consider a PT percolating at a 100 MeV temperature for precision: Requiring it to be slow and strong according to the above criteria implies a minimum duration on the μ s scale and a released equivalent mass density of roughly $10^{14} \text{ g cm}^{-3}$ respectively. The latter is approximately the mass density of a neutron star. Due to the sheer size of the expected energy flows, we can hence expect the emission of observable GWBs by such transitions.

*Dark bubble walls
only feel the DS*

*The transition
speed*

*Strong and slow
relative to what?*

4.3.1 Bubble collisions and sound waves

*The energy budget
of a PT*

The GW spectrum emitted in FOPTs is sourced by a set of intertwined physical processes, most importantly bubble collisions, as well as sound waves and turbulence in the perturbed plasma. Each of these contributions can potentially dominate the GW spectrum, depending in particular on the speed v_w of the bubble wall and the strength α_{DS} of the PT. Since predictions for plasma turbulence as a GW source often requires especially long lattice simulations with small grid spacing [191], which are not yet understood well enough to make quantitative statements [178], we will only consider sound wave and bubble collision contributions in this thesis. The energy budget of a PT is conventionally parameterized by the sum $\kappa_\phi + \kappa_{sw} + \kappa_{therm} = 1$, where the first contribution quantifies the percentage of liberated vacuum energy going into gradient field energy in the bubble walls, the second term corresponds to bulk fluid motion and κ_{term} corresponds to an immediate increase of the thermal energy. In this thesis, we will work in the high- v_w approximation discussed in the appendix of ref. [192] in which $\kappa_{therm} \rightarrow 0$. We use the expression provided there to compute the efficiency κ_{sw} in dependence of α_{DS} . In section 6.A we will discuss the validity of the high- v_w approximation in the $U(1)'$ model.

*The envelope
approximation*

BUBBLE WALL CONTRIBUTIONS The spectrum of GWs emitted by bubble collisions was first derived analytically and then studied in simulations using the so-called envelope approximation [193–195], in which a thin sheet with large gradient field energy around the collided bubbles is assumed to give rise to the anisotropic stress necessary for emitting GWs. Since then several analytical works [196–200] have been performed in order to refine the original results for the contribution of bubble wall collisions. In this thesis we will use the parameterization used in the LISA 2015 review [27] based on ref. [198],

$$h^2\Omega_{bw}(f) = \mathcal{R}h^2 \tilde{\Omega}_{bw} \left(\frac{\kappa_\phi \alpha}{\alpha + 1} \right)^2 \left(\frac{\beta}{H} \right)^{-2} s_{bw}(f/f_{p,bw})$$

with $s_{bw}(x) = \frac{3.8x^{2.8}}{1 + 2.8x^{3.8}}$ and $f_{p,bw}^{\text{em}}/\beta = 0.23$. (4.32)

The amplitude of the signal at the time of production is fixed by $\tilde{\Omega}_{bw} = 0.11v_w^3 / (0.42 + v_w^2) = 0.077$ for $v_w \rightarrow 1$. Assuming entropy conservation between the spectrum's emission and today, the amplitude redshift $\mathcal{R}h^2$ follows the previously derived eq. (2.75) and a given mode's frequency f today is related to an emitted frequency f^{em} through eq. (2.76) with $x_k/(2\pi)$ corresponding to the above expression for the spectrum's peak frequency at the time of emission $f_{p,bw}^{\text{em}}$. We see that the IR slope of the spectrum s_{bw} increases with $f^{2.8}$ (which is close to the causality IR tail with f^3 [86, 201–203]), whereas the UV slope decreases slowly with f^{-1} .

*The bulk-flow
model*

Numerical simulations of bubble collisions lately challenged the envelope approximation: The peak frequency showed to be a factor 5 higher than

previously expected and that the UV slope should fall off more quickly with $f^{-1.5}$ [204]. In ref. [205] it was further shown that both the IR and UV slopes depend on the finite thickness of the wall profile, confirming previous analytical works [200, 206] referred to as the bulk-flow model. A concise comparison of the different parameterizations of the bubble wall spectrum can be found in chapter 6 of ref. [207].

In a given particle physics model featuring a FOPT the contribution from bubble wall collisions is, however, often expected to be subdominant compared to the contribution from sound shell collisions since $\kappa_\phi \ll \kappa_{\text{sw}}$. Only in the case of bubble walls which run away, i.e. if the bubble walls accelerate up to the point of their collision, the dominant contribution to the total GW signal is expected to arise from bubble collisions [208].

Sound waves typically dominate

SOUND WAVE CONTRIBUTIONS The sound wave contributions to the total GWB spectrum were first studied in the early 90s [193]. Today, hydrodynamical simulations [180, 209–212] and corresponding analytical investigations [213, 214] referred to as the sound shell model sustain the parameterization of the GWB signal employed in the LISA 2019 review [178, 212], which we will also use in this thesis. It reads

$$h^2 \Omega_{\text{sw}}(f) = \mathcal{R} h^2 \tilde{\Omega}_{\text{sw}} \left(\frac{\kappa_{\text{sw}} \alpha}{\alpha + 1} \right)^2 \left(\frac{\beta}{H} \right)^{-1} \mathcal{Y}_{\text{sh}} s_{\text{sw}}(f/f_{\text{p,sw}})$$

with $s_{\text{sw}}(x) = x^3 \left(\frac{7}{4 + 3x^2} \right)^{7/2}$, $f_{\text{p,sw}}^{\text{em}}/\beta = 0.53$

and $\mathcal{Y}_{\text{sh}} = \min[1, \tau_{\text{sh}} H] \simeq \min \left[1, \frac{3.38}{\beta/H} \sqrt{\frac{1+\alpha}{\kappa_{\text{sw}} \alpha}} \right]$. (4.33)

The LISA sound wave template

The normalization of the signal is given by $\tilde{\Omega}_{\text{sw}} = 3 \times 0.012 \times 0.687 \times (8\pi)^{1/3} = 0.072$, where the first two factors were identified in ref. [212], the third one arises from the normalization of the spectrum $\int s_{\text{sw}} d\log x = 1/0.687$ and the fourth one converts the estimated mean bubble separations to β/H . Note that the initial factor 3 was erroneously missing in LISA 2019 review [178]. It should be noted that this expression can lead to a slight overestimation of the GW signal for sources being active for about a Hubble time. As the onset of turbulence is however an unsolved and model-dependent question, we conservatively stick to the shown expression for \mathcal{Y}_{sh} despite the developments shown in ref. [215].

Depending on the strength α_{DS} and the bubble wall velocity v_w , either a shock wave forms in front or behind the bubble walls. These cases correspond to sub- and supersonic bubble wall velocities, referred to as deflagration and detonation fluid profiles, respectively. In this thesis, we will use the fitting formula for $v_w \rightarrow 1$ provided in the appendix of ref. [192] to calculate the efficiency κ_{sw} for converting vacuum energy to bulk fluid motion. The factor \mathcal{Y}_{sh} accounts for the lifetime τ_{sh} of the sound wave source, after which shocks form and non-linearities develop. Note that the sound wave contribution is only suppressed by a single power in β/H , whereas the bubble wall collision contributions

The energy budget for fluid motion

scales with $(\beta/H)^{-2}$. Typically the bubble wall contribution is therefore suppressed even for similar energy budgets κ_ϕ and κ_{sw} .

4.3.2 The dilution of GWs due to a dark sector decay

So far we did not specify any couplings to SM states. As noted above, sufficiently small couplings (for instance to the SM photon through kinetic mixing or to the SM Higgs field through mass mixing) do not change the dynamics of the PT as their effect on the effective potential can be neglected. A PT in a secluded DS, which cannot dispose of the liberated vacuum energy density, is subject to strong constraints due to the aforementioned ΔN_{eff} constraints (cf. eq. (2.30)). We investigate these constraints in detail in chapter 5. In previous studies it was realized that a decay of the DS can circumvent these constraints at the cost of potentially changing the expansion history of the universe, especially if the DS is hotter than the SM bath and decays out-of-equilibrium [4]. The effect of the resulting entropy injection into the SM bath is an extra redshift between the time of GWB emission and today with respect to the naive Λ CDM expansion history. We refer to this additional decrease of the GWB amplitude and peak frequency as dilution [70]. It is characterized by the dilution factor

$$D \equiv \frac{S_{\text{SM}}^0}{S_{\text{tot}}^{\text{p}}} , \quad (4.34)$$

The effect of an entropy injection on $\mathcal{R}h^2$ and the peak frequency

which relates the comoving entropy density of both the DS and SM bath at the time of the PT with the current comoving entropy density of SM states. In case the DS thermalizes with the SM particles through the decay of a relativistic DS state, $D \rightarrow 1$ and the effect on the redshift of the GW signal is negligible. If instead a large amount of entropy is produced and injected into the SM bath, the previous expression for the frequency redshift f/f^{em} and amplitude redshift $\mathcal{R}h^2$ need to be corrected to [4]

$$f_{\text{p},i} = \frac{17 \text{ nHz}}{D^{1/3}} \left(\frac{\beta}{H} \right) \left(\frac{T_{\text{p}}}{100 \text{ MeV}} \right) \left(\frac{f_{\text{p},i}^{\text{em}}}{\beta} \right) \left(\frac{g_{\text{tot}}^{\text{p}}}{100} \right)^{1/2} \left(\frac{100}{h_{\text{tot}}^{\text{p}}} \right)^{1/3} \quad (4.35a)$$

$$\text{and} \quad \mathcal{R}h^2 = \frac{1.7 \cdot 10^{-5}}{D^{4/3}} \left(\frac{g_{\text{tot}}^{\text{p}}}{100} \right) \left(\frac{100}{h_{\text{tot}}^{\text{p}}} \right)^{4/3} . \quad (4.35b)$$

The temperature T_{p} here corresponds to the SM bath at the time of percolation, at which also the dofs $g_{\text{tot}}^{\text{p}}$ and $h_{\text{tot}}^{\text{p}}$ of the combined bath of SM and DS particles are evaluated, cf. eqs. (2.20a) and (2.20b).

A decaying DS can dilute the GWs

5

DO PTAS OBSERVE A DARK SECTOR PHASE TRANSITION?

This chapter is based on the following publication:

- [1] T. Bringmann, P. F. Depta, T. Konstandin, K. Schmidt-Hoberg, and C. Tasillo, *Does NANOGrav observe a dark sector phase transition?*, *JCAP* **11** (2023) 053, [[2306.09411](#)]

Never tell me the odds!

— Han Solo

5.1 INTRODUCTION

In the last chapter we found that DSPTs can be the source of strong GWBs. The recent PTA observation discussed in chapter 3 now leads us straight to the central question of the present chapter: *Do pulsar timing arrays observe a dark sector phase transition?*

At the end of chapter 3 we have already identified the strongest contender for any alternative explanation of the novel signal’s origin: an astrophysical GWB emitted by inspiraling SMBHBs. In order to explain the observed signal’s amplitude, however, it was argued that the local SMBHB density would need to be higher by an $\mathcal{O}(10)$ factor compared to previous estimates [216–218], cf. fig. 3.4. The question of whether realistic astrophysical models could give rise to a sufficiently strong GWB hence remains the subject of an ongoing debate [149, 219–222].

In order to match the observed GW signal at nHz frequencies, the preferred temperature for a cosmological PT must be at the MeV-scale [106, 121], cf. fig. 2.8. As new physics at this energy scale is very strongly constrained by a large variety of direct experimental searches [223], this immediately implies that the associated new states should only couple very weakly to the SM. In other words, such a phase transition would have to take place in a more or less secluded DS—which could, in fact, also be directly related to the dark matter puzzle [224–226]. Importantly, even if the DS is only very weakly coupled to the SM, a DSPT could impact the successful predictions of BBN and the CMB due to the extra energy density that is present in the DS (conventionally parameterized as an effective number of new relativistic neutrino degrees of freedom ΔN_{eff} , see eq. (2.30)) [12, 74] or through the late decay of

Astrophysics or
cosmology?

Stable and
decaying DSS

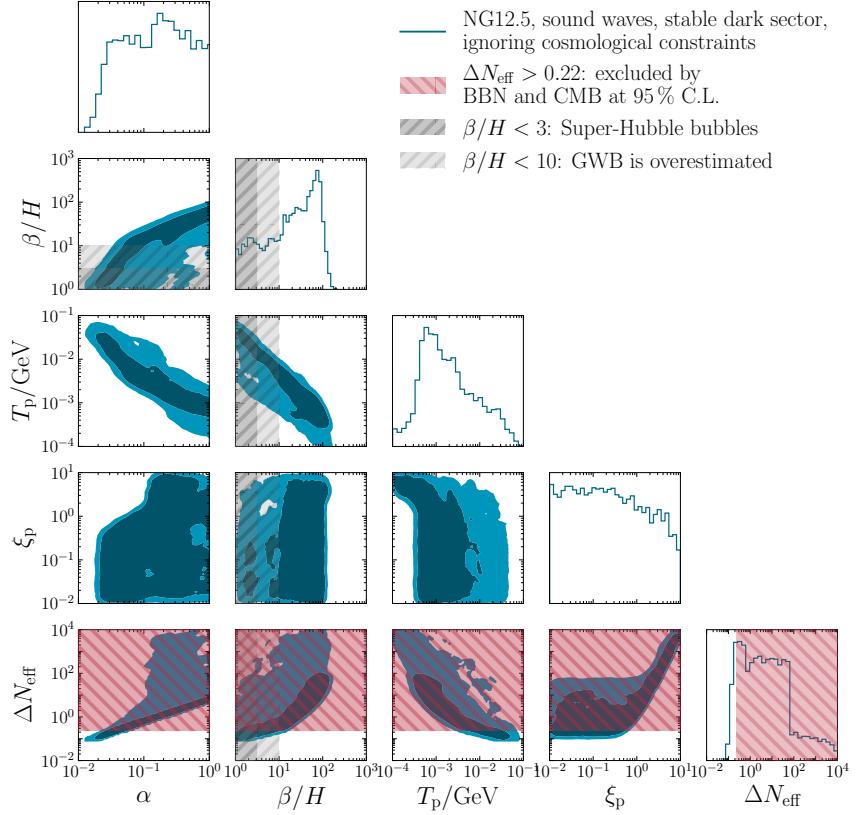


Figure 5.1: Triangle plot showing the 1σ and 2σ contours obtained by a naive fit (blue) of the NANOGrav 12.5yr data to a GW spectrum emitted in a DSPT, ignoring cosmological constraints. To illustrate the tension with BBN and CMB, the 95% C.L. excluded regions corresponding to $\Delta N_{\text{eff}} > 0.22$ are shaded in red [74], cf. the discussion in section 5.4. The regions in which super-Hubble bubbles ($\beta/H < 3$) and an overestimation of the GW background amplitude ($\beta/H < 10$) are expected are shaded in gray, see section 5.2 for further details.

the additional states [227–231]. In this chapter we will consider the two main generic possibilities, namely where the additional energy density

1. fully remains within the DS (“stable/secluded DS”), or
2. is subsequently injected into the SM sector (“decaying DS”).

It is worth stressing that these options simply refer to different regimes of the inter-sector coupling(s), and hence are *a-priori* equally viable from a phenomenological point of view. In both cases, in particular, we assume that those inter-sector couplings are sufficiently small for the DS to not thermalize with the SM heat bath. As a result the visible and dark sector will generally have different temperatures [25].

Fig. 5.1 illustrates in a nutshell the need to consistently combine cosmological and pulsar timing information when interpreting the NANOGrav

12.5yr results in terms of a DSPT.¹ The blue contours show the results of a naive fit of the DSPT parameters to NANOGrav data, *without* taking into account physically motivated priors on the rate β/H of the PT or cosmological constraints. We discuss the former in more detail in section 5.2, and the latter in section 5.4. Here, we simply want to demonstrate that these considerations (as indicated by gray and red shadings, respectively) will necessarily have a major impact on the naively inferred parameter space. One of our main results from a full statistical treatment, including information from cosmology, is indeed that an astrophysical explanation of the GW signal is much more credible than a GWB due to a PT from a *stable* DS for sub-horizon bubble sizes $R_*H_* < 1$ (corresponding to $\beta/H \gtrsim 3$, cf. eq. (5.49) in ref. [165]). When considering a DS that *decays* at pre-BBN temperatures, on the other hand, we find that the viable parameter space of DSPTs opens up: In this case, the NANOGrav data can be explained without violating BBN constraints, fitting the pulsar timing data as good as SMBHBs. For earlier works on cosmological constraints on PT interpretations of the NANOGrav results, see refs. [119, 232, 233].

This chapter is organized as follows: We start by discussing our parameterization of GWB spectra from DSPTs in section 5.2. In section 5.3 we continue with a detailed description of our statistical procedure to analyze PTA data, remarking also on pitfalls and limitations of simpler or more heuristic methods sometimes adopted in the literature. We describe the cosmological constraints on DS dynamics in section 5.4, and explain how to construct global likelihoods that simultaneously take into account pulsar timing and cosmological information. We present our results in section 5.5, before concluding in section 5.6. In three appendices, we provide further technical details about our analysis.

*A first glance at
the tension for
stable DSPTs*

*Outline of this
chapter*

5.2 COMPUTATION OF GW SPECTRA

In our analysis of the PT interpretation of the PTA data we use the parameterization of GWBs emitted through bubble collisions and sound waves provided in eqs. (4.32) and (4.33), respectively. In order to perform a data analysis that is as model-independent as possible, we further do not specify an underlying DS Lagrangian, but instead infer the PT parameters α , β/H , T_p and ξ_p occurring in the semi-analytical GW spectra in eqs. (4.32) and (4.33) through a fit with the PTA data. This allows us to understand the conditions a specific DS model needs to satisfy in order to explain the nHz signal. In the following paragraphs we want to briefly summarize the assumptions that go into our parameterization of the GWB.

*Model-independent
treatment*

¹ In the analysis presented in this chapter we used the now outdated NANOGrav 12.5yr data set instead of the latest 15yr data release. We comment on expected differences between the datasets in the context of our PT parameter inference in section 5.5.

We use $\xi_p \equiv T_{d,p}/T_p$ to denote the ratio of the DS temperature in the symmetric phase to that of the SM bath at the time of percolation. Assuming that the energy injection into the DS bath happens instantaneously after percolation, this means that ξ_p corresponds to the ratio just *before* the PT. To simplify our analysis, we also assume that the DS reheats instantaneously (cf. sec. 6.4.2 in the following chapter) and that the DS energy density after the transition is dominated by at least one relativistic particle species, such that the speed of sound is given by $c_s^{\text{DS}} = 1/\sqrt{3}$ throughout the transition.²

ξ_p : temperature ratio before the PT

Small β/H are problematic

UV tail of GWB spectra does not matter

Dilution D is negligible

Following the example of ref. [235], we demand $\beta/H > 3$ for successful percolation. This condition can also be interpreted as requiring that the bubbles have sub-horizon size during percolation, since $R_*H_* > 1$ implies $\beta/H > (8\pi)^{1/3} = 2.93$. Moreover, the time between the nucleation of the first bubbles to percolation is about $10/\beta$ (see, e.g., ref. [236]). As simulations neglect the expansion of the Universe during the PT, which suppresses the GW signal, spectra obtained from such simulations are therefore likely overestimated, or at least subject to sizable uncertainties, for $\beta/H < 10$.

Concerning the exact form of the GW spectra that we adopt here, let us mention that recent results seem to indicate sound shell decays leading to an f^{-3} rather than f^{-4} scaling in the UV [236]. This scaling has little impact on our results since very low phase transition temperatures are disfavored in our analysis, implying that the signal is not fitted by the (far) UV tail. Moreover, for wall velocities close to the speed of sound, the sound shell thickness becomes imprinted in the fluid motion [213, 214, 236, 237]. This leads to an additional knee in the power spectrum and an intermediate scaling x^1 . We also neglect this effect since we focus on wall velocities close to the speed of light.

In all scenarios studied in this chapter, we set the dilution factor appearing in eq. (4.35) to $D = 1$, corresponding to the assumption that there is no significant deviation from the standard cosmological redshift history. This is natural in our analysis of stable DSs, as the potential dilution gets sourced by an entropy injection, which can only happen for decaying DSs. A dilution factor $D > 1$ would correspond to a faster expansion than in radiation domination, e.g. if the PT were followed by an intermediate phase of early matter domination [4]. We checked that this dilution is always negligible also in the case of our decaying DS scenario, cf. appendix 5.A.2. We find that the assumption of radiation domination (corresponding to small deviations with respect to $D = 1$) holds within the parameter space favored by the data, thereby justifying this assumption a-posteriori (see below for a discussion and appendix 5.A.2 for further details).

² The authors of ref. [234] found that a slightly smaller speed of sound is expected in the broken phase of minimal DS models, which can lead to a sizable suppression of the GW signal for detonations. This would introduce a model dependence which we neglect in our analysis.

We further set $g_{\text{DS}}^{\text{p}} = h_{\text{DS}}^{\text{p}} = 1$. This is not a strong assumption, as the DS dofs always appear as prefactors of higher powers of ξ_{p} in the calculation. Since the experimentally preferred temperature ratios ξ_{p} are small, corresponding to cold DSPTs, the spectrum's dependence on the precise number of DS dofs is negligible. As discussed in section 5.4, this choice also does not influence the cosmological constraints that we impose.

*A specific choice
for DS dofs*

Finally, we treat bubble wall collisions and sound wave spectra, parameterized in eq. (4.32) and eq. (4.33), as two separate models. For bubble wall spectra, we set $\kappa_{\phi} = 1$ and consider only this contribution to the total GW spectrum. For sound waves, we compute $\kappa_{\text{sw}}(\alpha_{\text{DS}})$ using the high-velocity approximation from ref. [192], which yields $\kappa_{\text{sw}} \simeq 1$ for large α_{DS} in the case of non-runaway bubbles. Note that large α_{DS} are generally expected due to $\alpha_{\text{DS}} \propto \xi_{\text{p}}^{-4}$ in the limit of cold DSPTs $\xi_{\text{p}} \ll 1$ (cf. eq. (4.30)). Hence, the conversion efficiency of vacuum energy to bulk fluid motion is generally large, $\kappa_{\text{sw}} \rightarrow 1$, and ultra-relativistic bubble wall velocities $v_{\text{w}} \rightarrow 1$ are favored by the data.

*Fast dark bubble
walls*

In summary, we calculate the GWB spectrum based on the following set of DSPT parameters:

$$\{\alpha, \beta/H, T_{\text{p}}, \xi_{\text{p}}\}. \quad (5.1)$$

In fig. 5.2 we illustrate the generic influence of increasing α , β/H , and T_{p} on the sound wave and bubble wall collision spectra. The influence of ξ_{p} on the GWB spectrum turns out to be negligible in our analysis since cosmological constraints limit it to a sufficiently small value, cf. section 5.5. The spectra shown here correspond to the best-fit points of the analysis including cosmological constraints presented in section 5.5 (with a prior of $\beta/H > 1$).

*Four DSPT
parameters for the
GWB*

5.3 PULSAR TIMING ARRAY DATA ANALYSIS

In this section, we briefly review the commonly used methods for analyzing PTA data. We begin by discussing two approaches frequently found in the literature that are used to fit arbitrary GWBs to PTA data. Following that, we explain in detail why model comparisons based on global fits require a more rigorous analysis, specifically using the full PTA likelihood as implemented in NANOGrav's code `enterprise` [115, 116], cf. eq. (3.28).

*Our global fits
require `enterprise`*

The NANOGrav collaboration found that the CURN signal (see chapter 3) in their 12.5yr data set can be well-described by a power-law, cf. eqs. (2.71) and (3.29a),

$$h^2 \Omega_{\text{gw}}(f) = \frac{2\pi^2}{3H_{100}^2} f^2 h_{\text{c}}^2(f) = \frac{2\pi^2}{3H_{100}^2} A^2 \left(\frac{f}{1 \text{yr}^{-1}} \right)^{5-\gamma} \text{yr}^{-2}, \quad (5.2)$$

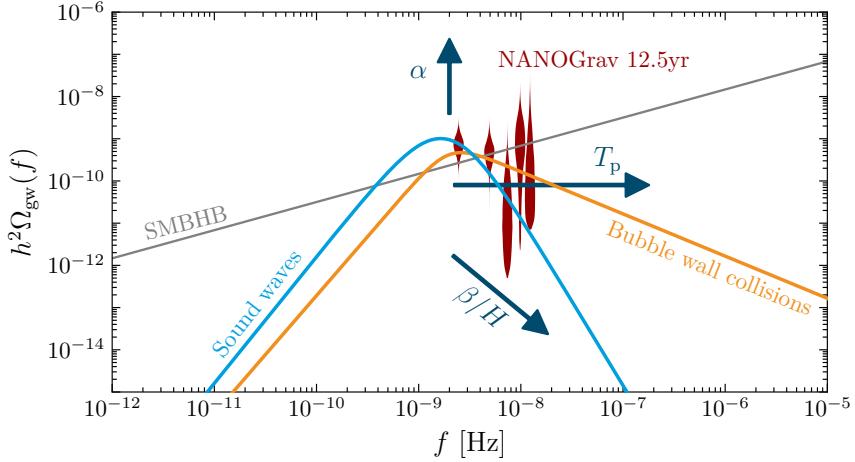


Figure 5.2: Plot showing the NANOGrav 12.5yr “violins” (red) [238], their standard explanation through a power-law spectrum from the inspiral of SMBHBs with $A_{\text{SMBHB}} = 1.53 \times 10^{-15}$ and $\gamma_{\text{SMBHB}} = 13/3$ (gray), as well as two PT spectra. A characteristic bubble wall collision spectrum is shown in orange. A sound wave-induced GWB spectrum is shown in blue. These spectra correspond to the best-fit parameter points found in section 5.5 (including cosmological constraints and demanding $\beta/H > 1$). The blue arrows indicate how an increase in the PT parameters α , T_p or β/H would shift the spectra.

where A and γ are the spectral amplitude and tilt, respectively. Keeping the tilt fixed to $\gamma = 13/3$ as expected for a GWB from inspiraling SMBHBs, see section 3.6, they reported a preferred signal amplitude $A \simeq 1.53 \cdot 10^{-15}$ [238].

An easy-to-implement possibility to fit *arbitrary* GWB spectra to PTA data is to reinterpret the posterior contours in the (A, γ) -plane produced by collaborations, cf. fig. 3.4, to also be valid for a GWB whose spectrum is close to a power-law in a certain frequency interval. This method has been used in many works [126–128, 136, 235, 239] that aim to explain the CURN signal by a GWB from cosmic origins rather than astrophysical sources. Since PTs result in GWBs with a broken power-law shape, this mapping to a specific combination of A and γ breaks down around the peak of the spectrum. While this method is often sufficient for estimating the approximate overall amplitude of the signal, it is not powerful enough for making a proper model comparison between different signal hypotheses.

Another tempting possibility to fit an arbitrary GWB spectrum to the CURN signal is to use the results of a free spectral analysis to the PTA data, which lead to the infamous “violins” reproduced in fig. 5.2. In that analysis the assumption of a power-law GWB is dropped in favor of free spectral amplitudes at the frequencies of integer multiples of $1/T_{\text{obs}}$. This approach then allows to directly fit spectra which can in principle deviate arbitrarily from a power-law (see, e.g., refs. [120, 240–242]). This

Mapping broken power-laws to A and γ is not sufficient

Using ceffyl would be tempting...

approach eventually led to NANOGrav’s own tool `ceffyl` [118] which promises a fast evaluation of the PTA likelihood, when marginalized over the pulsar-intrinsic red noise parameters, cf. eq. (3.30). A significant limitation of this approach is, however, that the posteriors for the respective frequency bins are calculated with a finite prior range of signal amplitudes. Adding to the fact that the tails of these posteriors are not necessarily sampled well enough, this implies that the violins cannot be used in any statistically meaningful way for signal amplitudes too far from their most likely values. For example, the finite extent of the violins shown in fig. 5.2 would strictly speaking imply that the ‘no-signal’ (i.e., the no common uncorrelated red noise (no-CURN)) hypothesis is excluded with arbitrary confidence—while in reality it is only disfavored by a Bayes factor of $\sim 10^4 - 10^5$ [238]!

Crucially for our analysis, cosmological constraints can force a potential DSPT to be so weak that the resulting GWB spectrum contributes only negligibly to the measured CURN. In such cases, the signal is fitted by fine-tuning the individual pulsar-intrinsic red noise amplitudes from eq. (3.29b). The methods discussed above implicitly rely on likelihoods that marginalize over these pulsar-intrinsic red noise parameters, making them unsuitable for our analysis. To properly account for correlations, a full evaluation of the likelihood is required. This approach to fitting cosmological GWB spectra using the full PTA likelihood has only been employed in a limited number of studies [121, 133], due to the high computational cost, and—to the best of our knowledge—never in a context that also incorporates additional constraining likelihoods, such as those from cosmology.

To interpret the NANOGrav data in terms of a DSPT we first construct a likelihood $\mathcal{L}_{\text{PTA}}(\boldsymbol{\theta}_{\text{PSR}}, \boldsymbol{\theta}_{\text{DS}})$ for fitting the timing residuals to a given set of pulsar-intrinsic red noise parameters $\boldsymbol{\theta}_{\text{PSR}}$ and a GW spectrum that depends on the DSPT parameters $\boldsymbol{\theta}_{\text{DS}}$. Each pulsar’s red noise is fitted by a power-law with an amplitude A_a and slope γ_a , cf. eq. (3.29b). To include constraints from cosmology on the available DS parameter space $\boldsymbol{\theta}_{\text{DS}}$, we further construct a likelihood $\mathcal{L}_{\text{cosmo}}(\boldsymbol{\theta}_{\text{DS}})$ in section 5.4. We multiply this likelihood with the PTA likelihood to obtain a global likelihood,

$$\mathcal{L}_{\text{glob}}(\boldsymbol{\theta}_{\text{PSR}}, \boldsymbol{\theta}_{\text{DS}}) = \mathcal{L}_{\text{PTA}}(\boldsymbol{\theta}_{\text{PSR}}, \boldsymbol{\theta}_{\text{DS}}) \times \mathcal{L}_{\text{cosmo}}(\boldsymbol{\theta}_{\text{DS}}). \quad (5.3)$$

In this analysis we concentrate on the NANOGrav 12.5 yr data [238], for which the full set of arrival time data, the pulsar white noise parameters as well as a tutorial on how to use these resources is publicly available [243]. In this data set, a total of 47 pulsars was taken into account, out of which we consider those that were observed for at least three years, as done in the original analysis [238]. From the remaining 45 pulsars, we treat the pulsar J1713+0747 as advertised in ref. [243] due to the probably mis-modeled noise process found in the dropout analysis [238]. The parameter space we evaluate therefore consists of 90 nuisance parameters $\boldsymbol{\theta}_{\text{PSR}} = \{A_a, \gamma_a\}$ for the pulsar-intrinsic red noise, adding to our four

... but leads to statistical fallacies for weak GWBs.

Constructing a global likelihood

Analysis choices for the 12.5yr data set

(five) DSPT model parameters $\boldsymbol{\theta}_{\text{DS}}$ for the case of a stable (decaying) DSPT. To evaluate this high-dimensional global likelihood in a numerically feasible way, we implement the DSPT spectra and $\mathcal{L}_{\text{cosmo}}(\boldsymbol{\theta}_{\text{DS}})$ into the codes `enterprise` and `enterprise_extensions` [115, 116]. To sample over the global likelihood we use `PTMCMC` [117]. The Markov chain-Monte Carlo (MCMC) chains are evaluated using `numpy` and `scipy` [244], and triangle plots are generated using `matplotlib` and a customized version of `ChainConsumer` [245]. For performing model comparisons, finally, we calculate Bayes factors using the product space method [246–249], which we briefly review in appendix 5.C.1.

5.4 COSMOLOGICAL CONSTRAINTS

The past decades of observational cosmology have provided a large amount of data which allow us to confidently reconstruct the cosmological evolution up to MeV-scale temperatures. Most notably these include observations of the CMB, both in terms of the spectral shape [250] and anisotropies [12], and the primordial light element abundances produced during BBN [39]. These observations are in very good agreement with the standard Λ CDM model and with each other [39], implying that any changes to Λ CDM at temperatures below a few MeV can have observational consequences and need to be checked for consistency with available CMB and BBN data [12, 74, 227, 228, 230, 231, 251].

For a PT to produce a strong GW signal a sizable amount of vacuum energy needs to be released in the transition, most of which is subsequently converted into DS energy density as only a small fraction ends up in the GWB. This additional energy density could change the well-tested cosmic expansion history, possibly even long after the transition itself. To understand whether PTAs may observe the remnants of such a PT we therefore need to include a cosmological likelihood $\mathcal{L}_{\text{cosmo}}$ when analyzing the PTA data. Specifically, we include information from the primordial light element abundances and CMB anisotropies into our analysis as described below.³

5.4.1 Stable dark sectors

If the entire DS energy density after the PT is contained in light degrees of freedom, this contributes an additional radiation energy density that can be described by a (potentially large) additional contribution ΔN_{eff} to the effective number of neutrinos $N_{\text{eff}} = N_{\text{eff}}^{\text{SM}} + \Delta N_{\text{eff}}$, where $N_{\text{eff}}^{\text{SM}} =$

³ Note that constraints from μ -distortions of the CMB photon spectrum [252] and curvature perturbations [253] are less relevant as they quickly lose sensitivity for transition temperatures above the MeV-scale. PBH formation due to first-order phase transitions [35, 254, 255] could offer novel probes, but turns out to be irrelevant for the phase transition strengths of interest in this analysis.

Precision cosmology at the MeV-scale

BBN and CMB constraints are most important

3.044 [61] is the SM prediction for Λ CDM cosmology.⁴ The effective number of neutrinos affects the predictions of BBN (see sec. 2.3.6) as well as the CMB power spectrum and is constrained by observations to $N_{\text{eff}} = 2.941 \pm 0.143$ [74]. This bound can be modeled by a Gaussian likelihood $\mathcal{L}_{\text{BBN+CMB}}(N_{\text{eff}})$, i.e.

$$\mathcal{L}_{\text{cosmo}}(\boldsymbol{\theta}_{\text{DS}}) = \mathcal{L}_{\text{BBN+CMB}}(N_{\text{eff}} = N_{\text{eff}}^{\text{SM}} + \Delta N_{\text{eff}}(\boldsymbol{\theta}_{\text{DS}})), \quad (5.4)$$

enabling us to implement the cosmological constraints for the case of a stable DS. The number of dofs in the DS does not have any relevant impact on cosmological constraints as the available latent heat would simply be distributed among the different species, leaving the total injected energy density unchanged. (Taking into account the energy density *before* the phase transition, a change in the number of dofs can be absorbed in the temperature ratio ξ_d which we treat as a free parameter in our scans anyway.) We therefore set $g_{\text{DS}}^{\text{P}} = 1$ in our analysis. For more details we refer to appendix 5.A.1.

Stable DSs are constrained through ΔN_{eff}

5.4.2 Decaying dark sectors

If there are additional small inter-sector couplings—which happens very naturally due to possible ‘portal’ couplings such as Higgs mixing for scalars or kinetic mixing for dark photons—the energy density injected into the DS will subsequently be transferred to the SM heat bath via decays of DS particles. In this case cosmological constraints in general depend on the lifetime, mass, and coupling structure of the decaying particles. As a simple concrete example and for minimality, we consider the DS after the PT to consist only of one bosonic dof ϕ decaying into photons or electrons with a lifetime τ_ϕ that we sample over. This is a natural setup in the context of a potential PT at MeV temperatures, as a light scalar dof with mass below the PT temperature is generally expected to exist in that case and, e.g., constraints from Higgs mixing would not be very severe.

Decaying DS constraints depend on the mediator’s mass and lifetime

Given that the NANOGrav data prefer an MeV-scale PT temperature, also the mass of ϕ is expected to be of this order, $m_\phi \simeq \text{MeV}$. In the example of Higgs mixing very short lifetimes τ_ϕ correspond to a sizable Higgs mixing angle θ which is constrained by a number of laboratory experiments to be $\theta \lesssim 10^{-4}$ for small masses [159, 256]. Translating this constraint for MeV-scale masses into the lifetime results in $\tau_\phi \gtrsim 10^{-2} \text{ s}$ whereas cosmological constraints roughly require $\tau_\phi \lesssim 10^{-1} \text{ s}$, so that some allowed region remains even in this scenario.⁵ As already indicated,

$\tau_\phi \lesssim 0.1 \text{ s}$

⁴ The assumption of a radiation-dominated DS is conservative in the sense that constraints only become tighter if the energy density instead starts to redshift as matter at some time after the phase transition. We also note that the contribution to ΔN_{eff} from the GWs themselves, for a GWB with peak amplitude $h^2 \Omega_{\text{GW}}^{\text{peak}} \lesssim 10^{-9}$ as required to explain the NANOGrav signal, is typically less than $\Delta N_{\text{eff}} \simeq 10^{-3}$, cf. ref. [37], and therefore irrelevant in the discussion of cosmological constraints.

⁵ Note that the case of a Higgs-mixed scalar is particularly constrained because of the Yukawa-suppressed couplings to electrons, implying a rather long lifetime. If the

the resulting cosmological constraints largely depend on the lifetime τ_ϕ of the particle, with only a very mild dependence on the mass as long as the energy density in ϕ is not very suppressed. MeV and smaller masses, on the other hand, generally lead to strong constraints for arbitrarily short lifetimes τ_ϕ due to thermalization by inverse decays. To simplify our analysis we therefore fix the mass of the decaying particle to $m_\phi = 5 \text{ MeV}$, noting that the results will be very similar for other choices of the mass in the MeV-range. To implement the cosmological constraints, we use a likelihood $\mathcal{L}_{\text{BBN+CMB}}$ incorporating the results from refs. [232, 251] (see appendix 5.A.2 for details).

5.5 RESULTS

We now present the results of our data analysis, based on global fits of the model setups described in the previous sections. We begin by examining PTs in stable DSs (section 5.5.1), followed by a discussion of a decaying DS that thermalizes with the visible sector some time after the PT (section 5.5.2). Finally, we compare with the SMBHB explanation of the signal in terms of the respective Bayes factors (section 5.5.3) and explore how later PTA data sets impact the DSPT interpretation.

5.5.1 Stable dark sector phase transitions

Let us first focus on GWs that are mainly emitted as a consequence of the bulk motion of the DS plasma after the PT, i.e. a GWB dominantly produced through sound waves. The full set of model parameters to describe such a scenario for a stable DSPT, as introduced in section 5.2, is given by

$$\{\alpha, \beta/H, T_p, \xi_p\}. \quad (5.5)$$

We sample over these input parameters with flat log priors, as well as over 90 nuisance parameters $\boldsymbol{\theta}_{\text{PSR}}$ for the pulsar-intrinsic red noise, based on the combined PTA and cosmological likelihood given in eq. (5.3). For a full overview over parameters and prior ranges, see also table 5.1 in appendix 5.C.6.

We show the resulting corner plot of posterior distributions for the four model parameters in fig. 5.3, to which we add the derived parameter ΔN_{eff} . Allowing for inverse time scales down to $\beta/H > 1$ (light blue) formally results in a good global fit of the pulsar timing residuals, as indicated by the compact ellipsoid-like posterior regions where the NANOGrav signal is explained by the GWB. Such small values of the transition rate would however suppress the GW spectrum w.r.t. the relevant state would be e.g. a dark photon, the allowed range of lifetimes would be significantly larger.

*First: Stable DS,
GWB from sound
waves*

*A seemingly good
fit for $\beta/H > 1$*

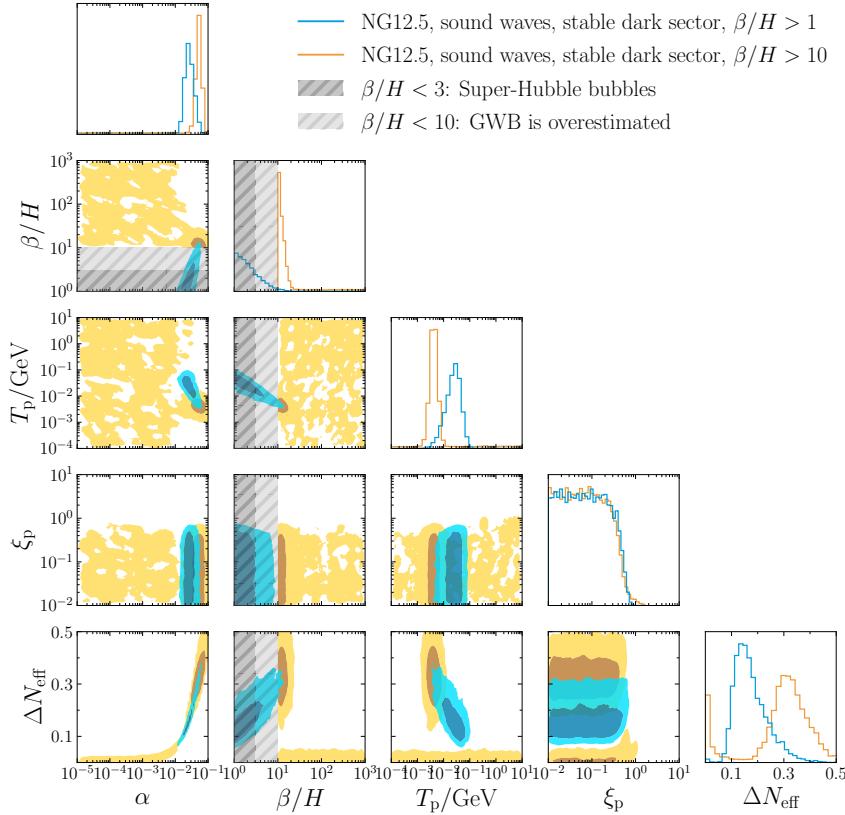


Figure 5.3: Comparison of the triangle plots for two MCMC chains with different priors on β/H , assuming a stable DS and a GWB generated through sound waves. The shaded regions mark 1σ and 2σ contours, respectively. Regions with $\beta/H < 3$ and $\beta/H < 10$, in which the phase transition cannot complete and the GW signal is overestimated respectively, are indicated with gray shadings.

commonly adopted parameterization, cf. the discussion in section 5.2, and, for $\beta/H \lesssim 3$, likely not even lead to successful percolation.

We therefore also show, in the same figure, the results of a fit with a more restrictive prior of $\beta/H > 10$ (orange). In this case, the best-fit region moves to somewhat larger values of α , but it also becomes apparent that there is no longer a single preferred region in the model parameter space: Instead, the combined data now shows a similar preference for a very weak DSPT-induced GW signal with correspondingly weak cosmological constraints, where the NANOGrav signal is not explained by the GWB but absorbed in the pulsar-intrinsic noise amplitudes A_a . This indicates that also the best-fit region no longer corresponds to an equally plausible interpretation of the combined data set.

For $\beta/H > 10$: No single preferred best fit region

The reason is that cosmology adds a constraint on ΔN_{eff} which effectively translates into a constraint on the PT strength α . In terms of fitting the NANOGrav signal, the required lower value of α can partially be

compensated by a lower value of β/H (see also fig. 5.2). And indeed, comparing the posterior distributions for β/H in fig. 5.3, we can see that β/H always sticks closely to the lower prior boundary—which is qualitatively different from the analysis without cosmological constraints, cf. fig. 5.1, where inverse timescales of $\beta/H = \mathcal{O}(10 - 100)$ are favored. Increasing the lower prior bound on β/H therefore directly induces a shift towards larger α . At the same time, a larger inverse timescale β/H also means smaller bubbles at the time of their collision, and hence a larger peak frequency in the spectrum (see again fig. 5.2). This is compensated for by a lower percolation temperature T_p , which by itself would lead to smaller peak frequencies. Finally, we can identify in fig. 5.3 an upper bound on the initial temperature ratio, which again is a direct consequence of the constraint on ΔN_{eff} . For $\xi_p \gtrsim 0.7$, in particular, eq. (5.11) would imply a violation of this constraint already for a single DS species that is not non-relativistic before the transition [25].

Overall, these effects push the posterior for ΔN_{eff} towards higher values. Since this is strongly punished by the cosmological part of the likelihood, however, that also explains the already mentioned appearance of a second preferred parameter region, characterized by a weak DSPT that corresponds to $\Delta N_{\text{eff}} \simeq 0$ (at the price of an unobservably small GW signal). We confirmed that this region of parameter space is indeed explained by fine-tuning the pulsar-intrinsic red noise parameters, rather than by a GWB, by directly comparing the marginalized posteriors of the nuisance parameters θ_{PSR} between the two chains depicted in [light blue](#) and [orange](#). Let us stress that this parameter range would have been impossible to reliably infer with simpler statistical methods, i.e. by re-fitting a power-law common red spectrum described by (A, γ) or by using the free-spectrum “violins” (see the discussion in section 5.3).

From the above discussion, one would expect the Bayes factor between the DSPT and the no-CURN hypothesis to decrease when increasing the lower prior bound on β/H , as higher and higher values of ΔN_{eff} are needed to explain the combined data. We confirm this expectation explicitly in fig. 5.4. Here, each of the colored dots corresponds to a separate MCMC chain that was employed to determine the Bayes factor between the two models by using the product space method explained in appendix 5.C.1. The corresponding lines serve as a cross-check for the prior dependence of the Bayes factors, see appendix 5.C.4 for further details. Yellow dots and lines refer to the case of a stable DSPT where the GWB production is dominated by sound waves; this corresponds to the same model as shown in fig. 5.3.

For comparison, we also show the case of a GWB that is mostly due to bubble wall collisions (red). Just from the point of view of the resulting spectrum, cf. fig. 5.2, one might expect that this could be a viable alternative. Compared to sound waves, however, bubble wall spectra receive an additional parametric suppression of $h^2 \Omega_{\text{GW}}^{\text{peak}}$ by a factor of $(\beta/H)^{-1}$. This induces the need for a larger α and hence an even stronger constraint on ΔN_{eff} , making the GWB hypothesis

Low ΔN_{eff} requires low α , β/H and ξ_p

For $\Delta N_{\text{eff}} \simeq 0$, signal is absorbed by $\{A_a, \gamma_a\}$

Evidence for DSPT depends on β/H prior

Bubble wall collisions suffer from additional β/H suppression

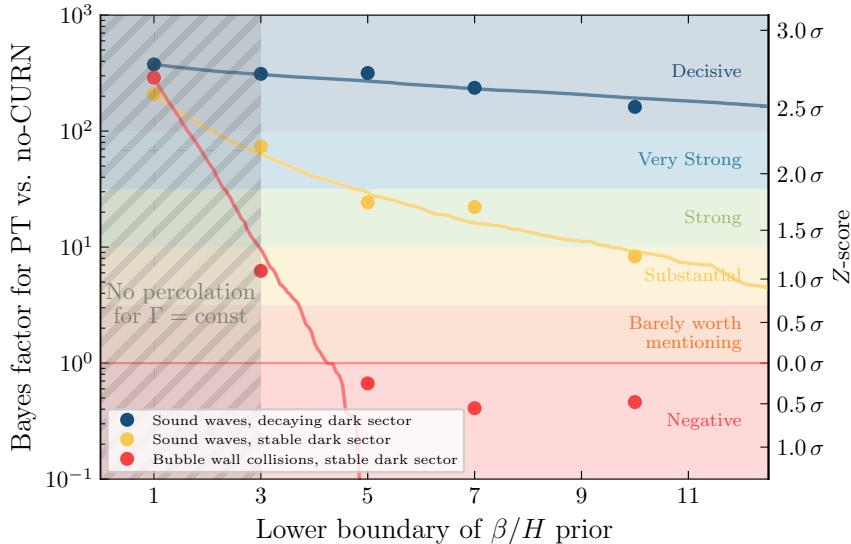


Figure 5.4: Bayes factor estimates of various DSPT models with respect to the no common red noise-hypothesis. Filled colored dots correspond to actually performed full model comparisons, while lines in the corresponding colors are derived from an *a posteriori* reduction of the prior range of chains with minimal β/H of 1 (for details, see appendix 5.C.4). We also indicate how to translate the Bayes factor to Jeffrey’s scale (colored shadings) as well as Z -scores from frequentist statistics (right y -axis, cf. appendix 5.C.3). In the gray shaded area ($\beta/H < 3$), a constant nucleation rate is not sufficient to drive percolation.

worse than the no-CURN assumption for $\beta/H \gtrsim 5$. To further illustrate these considerations, we refer to fig. 5.6 in appendix 5.B, showing the posterior distribution of the bubble wall spectra for different prior choices on β/H . Note also that neither fig. 5.3 nor fig. 5.4 includes the expected suppression in the GWB spectra for $\beta/H \lesssim 10$, which would further decrease our Bayes factor estimates.

Overall we therefore come to the conclusion that a stable DSPT can hardly compete with the alternative SMBHB explanation of the PTA timing residuals, once one takes into account cosmological constraints from BBN and CMB. For $\beta/H > 10$, in particular, the Bayes factor between a DSPT explanation and the no-CURN hypothesis is only $\mathcal{O}(10)$ even in the favorable case of sound wave-induced GWBs—much smaller than the factor of $\sim 10^{4.5}$ that is claimed for a GWB from SMBHBs [238].

Stable DSPTs are in strong tension with BBN and CMB

5.5.2 Decaying dark sector phase transitions

We next consider a DS that couples sufficiently strongly to ordinary matter such that it can decay after the PT. A decay long before BBN,

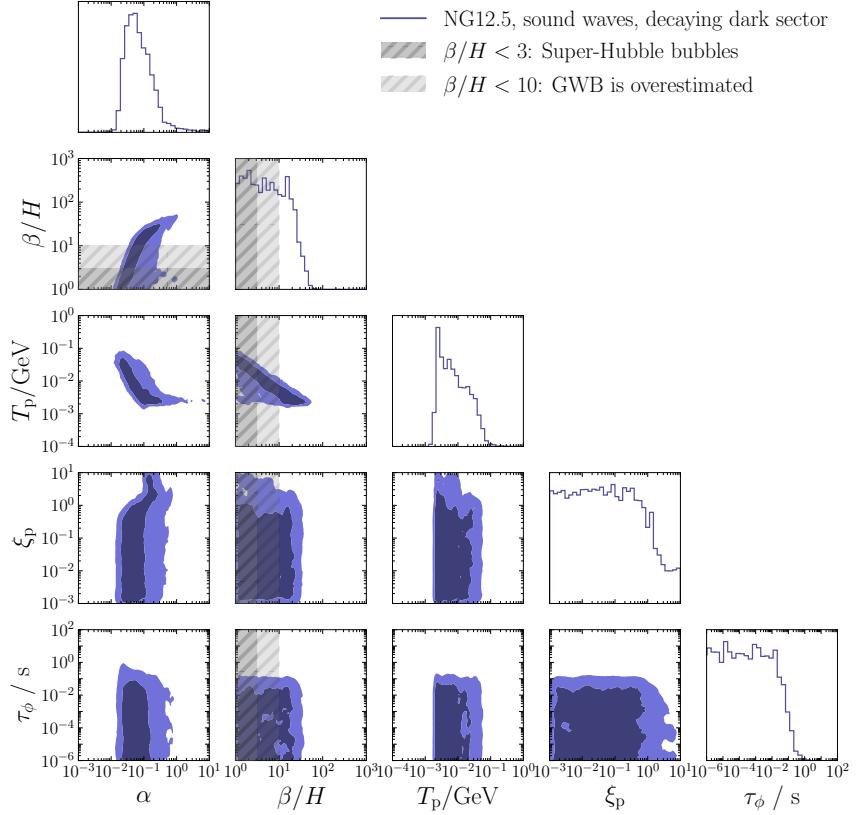


Figure 5.5: Posterior distributions of model parameters in the decaying dark sector scenario.

Can the DS decay before BBN?

in particular, is not subject to the cosmological constraints that played such a decisive role for the case of a stable DSPT. On the other hand, a PT happening too early would produce a GWB at too high frequencies to be compatible with the NANOGrav signal. It is therefore a non-trivial question whether any relevant parameter space remains where PTA and cosmological data are indeed compatible. For concreteness, we assume the decay of a dark Higgs boson as detailed in section 5.4 that decays with a lifetime τ_ϕ , such that our model parameters read

$$\{\alpha, \beta/H, T_p, \xi_p, \tau_\phi\}. \quad (5.6)$$

For the dark Higgs lifetime we adopt a log prior ranging from 10^{-6} s to 10^2 s; the remaining parameters we treat as in the previous section (with $\beta/H > 1$).

The decay requires $\tau_\phi \lesssim 0.1$ s and $T_p \gtrsim 2$ MeV

We show the resulting triangle plot for this model in fig. 5.5. In a nutshell, we find that the GW spectrum observed by NANOGrav can be explained as long as $\tau_\phi \lesssim 0.1$ s and $T_p \gtrsim 2$ MeV. Larger lifetimes, corresponding to temperatures smaller than 2 MeV, are strongly constrained as the decays occur after the onset of BBN or neutrino decoupling. Such decays change the time-temperature relation of the SM heat bath and alter the

ratio of the neutrino and photon temperatures, leading to a *negative* contribution to ΔN_{eff} . If the percolation temperature T_p drops to values close to the temperature of neutrino decoupling, strong constraints arise independent of the lifetime τ_ϕ . Note that we implemented the results from ref. [232] for simplicity as a sharp cut-off enforcing $T_p > 2 \text{ MeV}$ in our likelihood, cf. appendix 5.A.2. Adopting a more accurate likelihood would result in a smoother transition of the posterior in the range $T_p \approx 1 - 2 \text{ MeV}$, the main effect being a slight increase of the maximal possible value of β/H .

Compared to our analysis of stable DSPTs, the posterior for the inverse timescale β/H is relatively flat up to $\beta/H \sim 30$, implying a very limited prior dependence. The underlying reason for this is the possibility of dumping the liberated energy density into the SM photon bath before the neutrino decoupling at around 2 MeV, thereby evading cosmological constraints and hence opening up for large PT strengths $\alpha \gtrsim 0.1$ to fit the GW signal even for $\beta/H \gtrsim 10$. This however only works up to the point when β/H becomes so large that its effect on the peak frequency can no longer be compensated for by a correspondingly lower percolation temperature, cf. fig. 5.2.

In fig. 5.4 we also indicate the Bayes factor for the decaying DSPT scenario (blue). As anticipated, the prior dependence on β/H is much less severe than in the scenarios discussed previously. In particular, this shows that a GWB from a decaying DSPT is a viable explanation of the observed signal even for $\beta/H > 10$. Quantitatively, the model evidence is a factor of ~ 200 larger than that of the no-CURN hypothesis, corresponding to 2.5σ or a “decisive” evidence on Jeffrey’s scale.

The DSPT cannot be arbitrarily strong

“Decisive” evidence for decaying DSPT

5.5.3 Comparison with SMBHBs and later data sets

Let us next address in more detail the question of how a DSPT interpretation of the signal compares to alternative GWB hypotheses, in particular the leading astrophysical explanation of an SMBHB-induced GWB, and how future data will help to further distinguish these two.

We start by pointing out that the DSPT spectra actually fit the GW spectrum in the NANOGrav data slightly *better* than the SMBHB spectra. Naively, this is already expected from fig. 5.2, and we demonstrate this in more detail in appendix 5.C.5. Nevertheless the maximal Bayes factor (with respect to no-CURN) that we find is only about $10^{2.5}$, significantly smaller than the $\sim 10^{4.5}$ typically quoted for SMBHBs [238]. We checked explicitly that the reason is *not* connected to the goodness-of-fit, but entirely due to prior volume effects: Bayesian statistics dutifully renders the simple SMBHB explanation of the data more credible than the apparently more complicated DSPT model.

A better fit despite smaller Bayes factors

However, it is important to remember that the amplitude A_{SMBHB} of the astrophysical signal is not necessarily a single fundamental parameter,

*An equal footing
for the SMBHB
and PT priors?*

as assumed in deriving a Bayes factor of $10^{4.5}$. Instead, it is derived from more complex astrophysical models, involving parameters such as merger timescales, and the mass, redshift, and spatial distributions of the SMBHBs. A full Bayesian analysis should thus also consider constraints on these fundamental parameters, as e.g. done in ref. [216] without fitting the NANOGrav data. This would decrease the formal evidence for the SMBHBs interpretation because (*i*) the prior volume increases due to the intrinsic parameters that the amplitude depends on, and (*ii*) astrophysical models predict amplitudes that are smaller than those inferred from the NANOGrav data [216–218, 257]. Analogously, of course, the DSPT parameters α , β/H , T_p , ξ_p , and τ_ϕ are in reality derived quantities from a given SM extension whose independent parameters are masses and couplings. Evaluating specific models where these underlying parameters are known would be interesting and clearly deserves further study, but is not the aim of our more model-independent analysis.

Shortly after our calculations were published [1], the NANOGrav collaboration updated their data set and announced strong evidence in favor of the HD curve [102] during an internationally acclaimed press conference (see chapter 3). While the Bayes factor supporting a GW origin of the common red signal reported by this collaboration and other PTAs significantly increased, the relative odds between the SMBHBs explanation and an alternative DSPT interpretation are expected to remain mostly unchanged. As discussed in section 3, the PTA response to a GWB factorizes into a pulsar correlation-dependent part, Γ_{ab} (corresponding to the HD curve), and a common red noise spectrum, which we correctly anticipated to be due to a GWB. Since the specific form of Γ_{ab} only slightly affects the common red noise spectrum, the reconstructed spectrum received only minor, though interesting, updates.

*Evidence in favor
of the HD curve*

*Slightly stronger
signals with
positive spectral
slope*

Intriguingly, the favored spectral amplitude (for a fixed slope of $\gamma_{\text{SMBHB}} = 13/3$) increased slightly in the new NANOGrav 15yr data set, from $A_{13/3}^{12.5\text{yr}} = 1.5 \cdot 10^{-15}$ to $A_{13/3}^{15\text{yr}} = 2.4 \cdot 10^{-15}$. This suggests a slightly worse fit with SMBHB models, which generally predict lower signal amplitudes [7, 149, 238]. It is worth noting that the posterior distribution of the signal in the (A, γ) plane not only shifted toward stronger signals but also toward weaker spectral tilts. In the 12.5yr data, the posterior centered around $\gamma_{12.5\text{yr}} = 5.5_{-1.7}^{+1.3}$ (median with 90% credible interval), corresponding to an almost scale-invariant power-law in the $h^2\Omega_{\text{gw}}(f)$ parameterization (see fig. 5.2). However, in the 15yr data, the preferred slope decreased to $\gamma_{15\text{yr}} = 3.2_{-0.6}^{+0.6}$. As a result, the spectral slope remained distinct from the naively expected $\gamma_{\text{SMBHB}} = 13/3 = 4.33$ for a GW-driven inspiral of SMBHBs, but interestingly, the relative sign of the slope flipped. The preferred GWB spectrum now features a positive slope $h^2\Omega_{\text{gw}}(f) \propto f^{5-\gamma} = f^{1.8 \pm 0.6}$ instead of a flat plateau shape, as shown in fig. 2.8.

In the data analysis presented above, we found that our fit favors spectra peaking around 3 nHz. This can be understood by considering the small relative uncertainty in the first two Fourier modes of the 12.5yr data set,

as shown in fig. 5.2. Given the new observational preference for a positive spectral tilt in the NANOGrav 15yr data set—consistent with the other PTA data sets announced on the same day [8, 9] and the latest IPTA data release [257]—the peak of a DSPT signal is expected to shift toward somewhat higher peak frequencies and amplitudes. Consequently, the contours shown in the corner plots in figs. 5.1, 5.3 and 5.5 are expected to shift toward higher percolation temperatures, around $T_p \simeq 100$ MeV. Preliminary calculations and a comparison with ref. [106] indeed support this argument.

Higher temperatures generally face weaker cosmological constraints due to their association with larger mass scales and faster particle decays, which have a smaller impact on the primordial plasma during BBN and recombination. Nonetheless, the arguments presented earlier still hold: Cosmological constraints remain critical when studying stable DSPTs. The relative amount of liberated vacuum energy retained in the dark sector (parameterized as α , or equivalently as a contribution to ΔN_{eff} , see eq. (5.11)) is independent of temperature and can significantly alter the primordial element abundances produced during BBN and shift the peaks in the CMB anisotropy multi-pole spectrum, as demonstrated in fig. 5.1.

In the case of a decaying DSPT, the increase in the GW signal’s peak frequency allows for slightly higher (and thus less problematic) values of β/H : A higher $f_{\text{peak}} \propto T_p \times (\beta/H)$ (see eq. (4.33)) can be accommodated not only by an increased T_p but also by smaller bubbles, which correspond to larger β/H . These are viable only in the *decaying* DS scenario, as they are accompanied by an increase in the transition strength α in order to keep a sizable peak amplitude. These stronger transitions would contribute to the tension with ΔN_{eff} constraints in the stable DS scenario, but can be circumvented by DS decays. As a result, the constraints on the stable DS interpretation are expected to remain robust or get somewhat stronger (with some variation due to normalization factors proportional to the number of SM dofs in eq. (5.11)). Meanwhile, the relative evidence supporting a decaying DSPT is likely to increase slightly, thanks to the broader range of viable transition temperatures T_p and speeds β/H .

An interesting connection to recent developments in our understanding of GW spectra from PTs emerges from the evidence supporting a modest spectral slope of $\propto f^{1.8 \pm 0.6}$. In the sound shell model (see sec. 4.3.1) this presence of a double-broken power-law spectrum with an intermediate $\Omega_{\text{gw}}(f) \propto f$ plateau shape, rather than a distinct peak, is predicted [151, 258]. This flattening occurs due to several relevant length scales present in the phase transition, particularly for slow bubble wall velocities. A double-broken power-law could also hint at a PT generating the GWB through thick bubble wall collisions [200, 206, 259]. In this thesis, we assumed a single-broken power-law shape for the GW signal from PTs, as shown in eq. (4.33) and fig. 5.2, due to the expected high bubble wall velocities, $v_w \rightarrow 1$. It is conceivable that future PTA data sets could

New data sets favor
 $T_p \simeq 100$ MeV

Stable DSPTs are still in tension.

Decays still save the fit.

The emergence of a flattening GW spectrum?

in fact reveal a distinct plateau, indicating interesting features in the underlying PT dynamics. As long as the GW spectral shape remains uncertain, particularly at frequencies above 10, nHz (see fig. 3.2), any evidence supporting a more complex PT explanation is, however, largely driven by prior assumptions [259].

The determination of whether PTAs detect a GWB emitted from inspiraling SMBHBs or if the signal is instead of cosmological origin depends not only on accurately identifying the spectral shape but also on the search for GWB anisotropies [260–262]. As detailed in section 2.4.6, the amplitude of an astrophysical background is expected to correlate strongly with the inhomogeneous matter distribution in the universe, while a cosmological background is likely to be highly isotropic, similar to the CMB. So far, no significant signs of anisotropy or discreteness in the GWB have been observed [15, 263–265]. Additionally, the polarization of the GWB is anticipated to provide valuable information once detected [266–272]. The challenge remains to determine how to incorporate these additional sources of information into the global fit framework presented here.

Finally, it is important to emphasize that not only the evidence for a nHz GWB is likely to strengthen in the future [273]; the competing cosmological constraints on PTs are also expected to improve. Upcoming experiments like the Simons Observatory and CMB-S4 measurements, in combination with large-scale structure surveys, will be capable of tightening the limits on ΔN_{eff} by about an order of magnitude [274–276]. This will increase the tension on the stable DS explanation we have explored and further reduce the parameter space for a decaying DSPT. Additionally, measurements of CMB spectral distortions with PIXIE [277] will provide complementary constraints [252], which will be relevant even in scenarios where a decaying DS avoids constraints on ΔN_{eff} . In fact, confirming spectral distortions in the CMB could potentially offer *supporting* evidence for such a scenario. Therefore, it will remain crucial to incorporate all relevant cosmological observables and new PTA data to ultimately determine the most probable origin of the observed GWB signal.

The case of GWB anisotropies

Future probes of precision cosmology

5.6 CONCLUSIONS

We investigated the appealing possibility that the GW spectrum observed by the NANOGrav collaboration in their 12.5yr data set [238] is due to a DSPT just before the onset of BBN. For the first time, we performed a global analysis on PTA data from a GWB including constraints from BBN and the CMB anisotropies.

We found that a DS undergoing a PT can in principle explain the measured signal with a goodness-of-fit that is comparable to (or even

better than) that of the standard astrophysical explanation in terms of a stochastic GWB from SMBHBs. However, if one accounts for

1. the changes in the early element abundances that the energy density released during the PT would induce,
2. the impact on the CMB anisotropies through a contribution to ΔN_{eff} , and
3. possible issues for transitions with $\beta/H < 10$, connected to percolation and an overestimation of the produced GWB,

the possibility of a *stable* DSPT no longer gives a good fit to all available data. Fig. 5.1 provides an intuitive illustration of this tension, by directly confronting the above constraints with the results of a naive DSPT analysis of the NANOGrav data that ignores them. Fully including all relevant constraints in a global fit, the available parameter space is indeed significantly reduced, cf. fig. 5.3.

On the other hand, there is no intrinsic reason why a DS should be stable on cosmological timescales. In particular, tiny interactions with the visible sector (e.g. through small portal couplings [278]) could well lead to a decay before neutrino decoupling at $T_p \gtrsim 2 \text{ MeV}$. We find that such a *decaying* DSPT scenario remains a compelling alternative to the more conventional SMBHB hypothesis for lifetimes $\tau_\phi \lesssim 0.1 \text{ s}$, cf. fig. 5.5. We arrived at this conclusion by further taking into account constraints on electromagnetic energy injection from decaying dark scalars [251] and on the reheating of the photon bath after a PT [232]. Compared to the no-GWB hypothesis, we find a Bayes factor that indicates a decisive evidence for the DSPT interpretation even for a prior of $\beta/H > 10$ on the transition rate. The currently maximal value of this quantity that is compatible with the data, $\beta/H \lesssim 50$, still indicates the need for a relatively slow transition; further model-dependent research will be needed to investigate how this can be implemented in a given SM extension.

We also studied the effect of prior choices on the absolute scale of Bayes factors, finding that prior volume effects are highly relevant when comparing SMBHB and DSPT explanations of the NANOGrav data. The SMBHB interpretation, in particular, seemingly only requires one parameter to fit the signal, namely the amplitude A . We however argue that A should rather be treated as a derived quantity that depends on several intrinsic, independently measured astrophysical quantities [216]. This would reduce the difference between the Bayes factors above 10^4 for the SMBHB explanation [238] and the Bayes factors of $\mathcal{O}(10^2)$ that we find for the decaying DSPT interpretation.

We remain excited about the pending third IPTA data release and searches for possible signs of anisotropy. While we do not expect a definite answer concerning the origin of the signal within the coming months, we are confident that our central results are robust also when considering the latest NANOGrav 15yr data set. Further, we are opti-

The stable DSPT explanation is in tension

Decaying DSPTs are a compelling alternative to SMBHBs

The need for an equal footing of priors

Global fits will be key to disentangling GWBs

mistic that additional PTA data as well as complementary constraints from cosmology [252] will help to disambiguate between different models in the future. With additional data it will be ever more crucial to fully include complementary constraints such as from BBN and CMB when assessing different signal models. In this chapter we have made a first step in this direction, thereby contributing to moving the realm of testable cosmology to pre-BBN times.

5.A DETAILS ON THE COSMOLOGICAL LIKELIHOOD

In this appendix we provide details on the cosmological likelihoods that we adopt in our analysis and discuss how to compute the relevant quantities that enter, including a mapping to previously published results.

5.A.1 Stable dark sector

For the case of a stable DSPT we conservatively assume that the energy density in the DS after the transition purely consists of radiation (cf. footnote 4). We focus on transitions before the onset of BBN and neutrino decoupling, such that the DS energy density can entirely be described by a simple additional contribution ΔN_{eff} to the effective number of neutrinos N_{eff} . This assumption is validated a-posteriori by the results of our MCMC chains, which indicate that the phase transition indeed happens sufficiently early.

Reproducing the ΔN_{eff} limits from ref. [232]

At the time of percolation, the DS energy density is thus given by

$$\rho_{\text{DS}}^{\text{p}} = \frac{\pi^2}{30} g_{\text{DS}}^{\text{p}} (\xi_{\text{p}} T_{\text{p}})^4 = \rho_{\text{DS}}^{\text{reh}} - \Delta V_{\text{eff}}, \quad (5.7)$$

where the index reh (p) indicates a point in time immediately after (before) the DS reheating and we assumed instantaneous reheating in the second step. Since $\Delta\theta_d/4 \simeq \Delta V_{\text{eff}}$ for sufficiently strong transitions, we find

$$\frac{\rho_{\text{DS}}^{\text{reh}}}{\rho_{\text{SM}}^{\text{p}}} \stackrel{(4.29)}{=} \alpha + (1 + \alpha) \frac{\rho_{\text{DS}}^{\text{p}}}{\rho_{\text{SM}}^{\text{p}}} = \alpha + (1 + \alpha) \frac{g_{\text{DS}}^{\text{p}}}{g_{\text{SM}}^{\text{p}}} \xi_{\text{p}}^4. \quad (5.8)$$

The total radiation energy density can be quantified in terms of the effective number of relativistic neutrino species,

$$N_{\text{eff}} = \frac{8}{7} \left(\frac{11}{4} \right)^{4/3} \frac{\rho_{\text{DS}} + \rho_{\nu}}{\rho_{\gamma}}. \quad (5.9)$$

Subtracting the SM contribution $N_{\text{eff}}^{\text{SM}} = 3.044$ [61], we find the DS contribution as

$$\Delta N_{\text{eff}} = \frac{8}{7} \left(\frac{11}{4} \right)^{4/3} \frac{\rho_{\text{DS}}}{\rho_{\gamma}} \quad (5.10)$$

$$\simeq \frac{4}{7} \left(\frac{11}{4} \right)^{4/3} \left(\frac{3.93}{h_{\text{SM}}^{\text{p}}} \right)^{4/3} g_{\text{SM}}^{\text{p}} \times \left[\alpha + (1 + \alpha) \frac{g_{\text{DS}}}{g_{\text{SM}}^{\text{p}}} (\xi_{\text{p}})^4 \right], \quad (5.11)$$

where we used $\rho_{\text{DS}}^{\text{reh}} = \rho_{\text{SM}}^{\text{reh}}$, inserted the SM degrees of freedom today and assumed the DS degrees of freedom to remain constant, $g_{\text{DS}} = g_{\text{DS}}^{\text{p}}$. This reproduces eq. (1) from ref. [232].

$\mathcal{L}_{\text{cosmo}}$ for stable
DSPTs

Observationally, N_{eff} affects the predictions of BBN as well as the CMB power spectra. Combining data sets from the Planck satellite with observations of the primordial abundances of deuterium and helium-4, and marginalizing over the baryon-to-photon ratio $\eta = n_b/n_\gamma$, ref. [74] finds $N_{\text{eff}} = 2.941 \pm 0.143$. We approximate this by a Gaussian likelihood:

$$\begin{aligned} \mathcal{L}_{\text{cosmo}}(\alpha, \xi_p, T_p) &= \frac{1}{\mathcal{N} \sqrt{2\pi\sigma_{N_{\text{eff}}}^2}} \\ &\times \exp \left[-\frac{(\Delta N_{\text{eff}}(\alpha, \xi_p, T_p) + N_{\text{eff}}^{\text{SM}} - \mu_{N_{\text{eff}}})^2}{2\sigma_{N_{\text{eff}}}^2} \right], \end{aligned} \quad (5.12)$$

where $\mu_{N_{\text{eff}}} = 2.941$, $\sigma_{N_{\text{eff}}} = 0.143$, $N_{\text{eff}}^{\text{SM}} = 3.044$, and the normalization needed for the correct determination of Bayes factors is given by

$$\mathcal{N} = \frac{1}{\sqrt{2\pi\sigma_{N_{\text{eff}}}^2}} \exp \left[-\frac{(N_{\text{eff}}^{\text{SM}} - \mu_{N_{\text{eff}}})^2}{2\sigma_{N_{\text{eff}}}^2} \right]. \quad (5.13)$$

We note that in the above construction we do not have to take into account differences that are reported between one-sided ($\Delta N_{\text{eff}} \geq 0$) and two-sided ($\Delta N_{\text{eff}} \in \mathbb{R}$) limits on additional radiation degrees of freedom, since we directly use the likelihood in our calculations. In particular, one-sided limits on ΔN_{eff} simply result from an integration of the likelihood starting from $N_{\text{eff}}^{\text{SM}}$ rather than smaller values (e.g. ref. [74]).

5.A.2 Decaying dark sector

Mapping ref. [251]
to the decaying DS
scenario

Constraints from BBN and CMB on a decaying particle with MeV-scale mass have been calculated in ref. [251] for a general setup. Here the initial conditions for ϕ , at some SM temperature $T_{\text{cd}} = 10 \text{ GeV}$,⁶ were taken to be a Bose-Einstein distribution with temperature $T_{\text{d,cd}} = \xi_{\text{cd}} T_{\text{cd}}$ and zero chemical potential, i.e.

$$f_\phi^{\text{cd}}(p) = \left[\exp \left(\frac{\sqrt{p^2 + m_\phi^2}}{\xi_{\text{cd}} T_{\text{cd}}} \right) - 1 \right]^{-1}, \quad (5.14)$$

with corresponding number density $n_\phi^{\text{cd}} = \int \frac{d^3 p}{(2\pi)^3} f_\phi^{\text{cd}}(p)$. To use the results and constraints from ref. [251] we therefore need to map to this scenario.

Sufficiently strong intra-sector couplings lead to a self-thermalisation of the DS quickly after the phase transition. If these processes are ϕ -number violating, the chemical potential of ϕ will generally vanish. The

⁶ The subscript cd refers to chemical decoupling, having a setup of thermal production of dark matter by freeze-out (chemical decoupling) in a DS in mind.

phase-space distribution function after the DS reheating can hence again be described by a Bose-Einstein distribution function

$$f_\phi^{\text{reh}}(p) = \left[\exp\left(\frac{\sqrt{p^2 + m_\phi^2}}{\xi_{\text{reh}} T_{\text{reh}}}\right) - 1 \right]^{-1}, \quad (5.15)$$

with corresponding number and energy densities

$$n_\phi^{\text{reh}} = \int \frac{d^3 p}{(2\pi)^3} f_\phi^{\text{reh}}(p), \quad (5.16)$$

$$\rho_\phi^{\text{reh}} = \int \frac{d^3 p}{(2\pi)^3} \sqrt{p^2 + m_\phi^2} f_\phi^{\text{reh}}(p). \quad (5.17)$$

To compute the DS energy density we again assume instantaneous reheating, cf. eq. (5.7), and compute

$$\rho_\phi^{\text{reh}} = \rho_{\text{DS}}^{\text{reh}} \stackrel{(5.8)}{=} \frac{\pi^2}{30} g_{\text{SM}}^{\text{p}} T_{\text{p}}^4 \left[\alpha + (1 + \alpha) \frac{g_{\text{DS}}^{\text{p}}}{g_{\text{SM}}^{\text{p}}} \xi_{\text{p}}^4 \right]. \quad (5.18)$$

For fixed mass and lifetime, cosmological constraints are mostly driven by the comoving number density of a particle. We hence compute n_ϕ^{reh} from eq. (5.16), where we obtain ρ_ϕ^{reh} by numerically solving eq. (5.17) for $T_{\text{reh}} = \xi_{\text{reh}} T_{\text{p}}$. From this, we get

$$n_\phi^{\text{cd}} = n_\phi^{\text{reh}} \frac{s_{\text{SM}}^{\text{cd}}}{s_{\text{SM}}^{\text{reh}}} = n_\phi^{\text{reh}} \frac{h_{\text{SM}}^{\text{cd}}}{h_{\text{SM}}^{\text{reh}}} \left(\frac{T_{\text{cd}}}{T_{\text{reh}}} \right)^3, \quad (5.19)$$

which we in turn can solve numerically for the initial temperature ratio ξ_{cd} , from eq. (5.14), and thus map the case of a DSPT to the case computed in ref. [251]. We comment below on the validity of this mapping.

To construct a likelihood, we use the calculated primordial light element abundances, their theoretical errors from nuclear rate uncertainties, and N_{eff} underlying figure 6 (left) from ref. [251]. These are compared to the recommended values of the observed primordial light element abundances of deuterium $D/{}^1H^{\text{obs}} = (2.547 \pm 0.025) \times 10^{-5}$ and the mass fraction of helium-4 $\mathcal{Y}_p^{\text{obs}} = (2.45 \pm 0.03) \times 10^{-1}$ [39] as well as $N_{\text{eff}}^{\text{obs}} = 2.99 \pm 0.17$ from CMB observations alone [12]. Noting that BBN calculations strongly depend on the baryon-to-photon ratio η , and that N_{eff} is affected by the DS decays, we use the best-fit value of η for given N_{eff} from figure 26 (Planck TT, TE, EE+lowE+lensing+BAO) of ref. [12]. Due to the strong η dependence of $D/{}^1H$, in particular, we need to propagate the uncertainty in the determination of η , and the total observational error effectively becomes $\Delta_\eta(D/{}^1H)^{\text{obs}} = 0.035 \times 10^{-5}$ [251]. The total cosmological likelihood is given by a product of Gaussian likelihoods, including also the aforementioned constraint on N_{eff} , with total errors obtained by summing the observational and theoretical errors in quadrature.

*On combining
BBN and CMB
observables*

For small τ_ϕ , the dark sector becomes visible

In the calculations outlined above we started from eq. (5.18), crucially assuming that the DS and the SM only thermalize after reheating, i.e. that (inverse) decays of ϕ only become relevant after reheating. This holds as long as $\tau_\phi > t_{\text{reh}}$, where t_{reh} is related to the Hubble rate through $H(T_{\text{reh}}) \simeq 1/(2t_{\text{reh}})$. The two sectors thus have equal temperatures soon after reheating, $\xi_{\text{reh}} = 1$, but T_p is no longer equal to T_{reh} . We find $T_{\text{reh}} = T_{d,\text{reh}}$ by solving

$$\rho_{\text{SM}}^{\text{reh}} + \rho_\phi^{\text{reh}} = (1 + \alpha)(\rho_{\text{SM}}^p + \rho_{\text{DS}}^p) \quad (5.20)$$

for T_{reh} with $\xi_{\text{reh}} = 1$ in eq. (5.15). Given T_{reh} , ρ_ϕ^{reh} , and ξ_{reh} , we can then find n_ϕ^{reh} as before and map to the results of ref. [251] by using eq. (5.14). This however still assumes that there is no change in the comoving number density of ϕ , and the comoving SM entropy density, between $T_{cd} = 10 \text{ GeV}$ and T_{reh} , i.e. that (inverse) decays can be neglected before reheating. Strictly speaking, this assumption is not valid due to thermalization around the time $t \sim \tau_\phi < t_{\text{reh}}$. However, successful thermalization erases all knowledge of initial conditions, implying that our mapping becomes inaccurate only if the thermalization itself has observable consequences, i.e. if it occurs during BBN. In that case, also the reheating process would have to occur during BBN, making it necessary to include, e.g., also the effect of (re)heating the SM. This was studied in ref. [232], with the result that the phase transition strength is constrained as $\alpha_* \gtrsim 0.07$ if $T_p \lesssim 2 \text{ MeV}$ for reheating into photons (under the assumption of $\xi_p = 0$). To compare this to our case we note that in ref. [232] the transition strength is defined by $\alpha_* = \Delta\theta_d/(4\rho_{\text{SM}}^p)$ such that $\rho_{\text{SM}}^{\text{reh}} = (1 + \alpha_*) \rho_{\text{SM}}^p$ and we need to map

$$\alpha_* = \alpha + (1 + \alpha) \frac{g_{\text{DS}}^p}{g_{\text{SM}}^p} \xi_p^4 - \frac{\rho_\phi^{\text{reh}}}{\rho_{\text{SM}}^p}. \quad (5.21)$$

Include results of ref. [232] as hard cutoffs

For simplicity, we implement these results as a hard cut on α . Our total cosmological likelihood is thus given by

$$\begin{aligned} \mathcal{L}_{\text{cosmo}}(\tau_\phi, \alpha, \xi_p, T_p) \\ = \frac{1}{\mathcal{N}} \times \mathcal{L}_{\mathcal{Y}_p} \times \mathcal{L}_{D/^{1\text{H}}} \times \mathcal{L}_{N_{\text{eff}}} \\ \times \begin{cases} 1 & \text{for } \tau_\phi > t_{\text{reh}} \\ \theta(\max[0.07 - \alpha_*, T_p - 2 \text{ MeV}]) & \text{for } \tau_\phi < t_{\text{reh}} \end{cases}, \end{aligned} \quad (5.22)$$

where the normalization is $\mathcal{N} = \mathcal{L}_{\mathcal{Y}_p} \times \mathcal{L}_{D/^{1\text{H}}} \times \mathcal{L}_{N_{\text{eff}}}$ for standard ΛCDM cosmology with only the SM contributing to the energy density during BBN.

The likelihood given above accurately describes the relevant cosmological constraints on a decaying scalar in a quite model-independent way. We however needed to make some assumptions along the way, either to assure the numerical feasibility of our calculations or to keep the number of parameters describing the decaying dark sector scenario low to allow

for a straight-forward phenomenological interpretation of our results. We stress that these assumptions generally do not affect our conclusions that the NANOGrav signal can be explained well by a decaying DSPT as long as the energy from the DS is injected into the SM before the onset of BBN and neutrino decoupling (i.e. $\tau_\phi \lesssim 0.1$ s and $T_p \gtrsim 2$ MeV). For completeness, let us still discuss the effect of these assumptions in some more detail.

A first simplification is the choice of the specific value of $m_\phi = 5$ MeV for the mass, due to readily available data from ref. [251]. Generally, the dependence on the mass is expected to be very mild. Only when the mass is small enough ($m_\phi \lesssim 2$ MeV) for the decaying particle to thermalize with the SM heat bath as a relativistic particle for $\tau_\phi \lesssim 0.1$ s and act as an additional relativistic dof, this changes abruptly and arbitrarily small lifetimes can be constrained [251].

*Small dependence
on m_ϕ*

Next, we use results for decays into photons. The results of ref. [251] show that decays into electron-positron pairs give very similar constraints for $m_\phi > 2m_e$. For smaller masses, these decays are kinematically forbidden such that a corresponding coupling will, in fact, mostly lead to decays into two photons. Constraints on the *total lifetime* thus become largely identical to the case of tree-level decays into photons (whereas constraints on couplings would become suppressed by loops and additional SM couplings). There are of course further interesting portals that lead to a thermalization between the two sectors, e.g. the kinetic coupling of a dark photon field with the SM photon. Such couplings are beyond the scope of this chapter due to additional model-dependent constraints. Further model-dependent complications, not changing the qualitative picture, may arise in case the assumption of instantaneous reheating is not justified (or if thermalization within the DS is not sufficiently fast, due to smaller intra-sector couplings).

*Non-
electromagnetic DS
decays*

By redshifting the number density of the decaying particle as in eq. (5.19), furthermore, we neglected a possible change in the comoving number density due to decays and inverse decays between T_{cd} and T_{reh} . While our mapping to the results of ref. [232] redeems this shortcoming adequately, a full model-dependent study should calculate the whole cosmological evolution of the DS including all relevant energy transfers between the DS and the SM. In this way, also the dilution factor D entering the GWB spectrum would be calculated [4], which we simply set to 1. This assumption is valid as long as the decaying particle does not become non-relativistic for an extended period of time before its decays. We checked a-posteriori how large the dilution factor would be for the regions favoured by NANOGrav data, cf. fig. 5.5, using the `dilution` sub-package of `TransitionListener` from ref. [4]. We find that, within the 1σ contour, D deviates by at most 2 % from 1. The maximal dilution factor within the 2σ contour is $D \approx 2$. Hence, our choice of setting $D = 1$ is not only conservative but also well justified in the relevant parameter space.

*Dilution is
negligible*

Number-changing processes in the DS can become important

Finally, our assumption of a negligible chemical potential—valid for efficient number-changing processes—might be violated in some setups. In particular, if the decaying scalar is not completely relativistic after reheating and the scalar self-couplings are low, number-changing processes can cease to be relevant and a chemical potential can develop [279, 280]. Such a scenario would likely require a full numerical solution of the respective Boltzmann equations and is therefore beyond the scope of this chapter.

5.B POSTERIOR DISTRIBUTION OF GWB SPECTRA

Constraints can push the spectra below the “violins”

To further demonstrate the possible tension between cosmological constraints and the interpretation of the NANOGrav data in terms of a DSPT, we illustrate the posterior distribution of GWB spectra for different parameter scans in fig. 5.6.⁷ The orange curves show the distribution of bubble wall collision spectra from a DSPT with $\beta/H > 1$. For bubble wall collision spectra with $\beta/H > 10$ (yellow curves), the tension in the data becomes apparent through the low signal amplitudes. The spectra cannot explain the GW spectrum, as depicted by the red “violins”, due to the strong constraints from ΔN_{eff} . When considering instead a decaying DS (violet), a PT with $\beta/H > 10$ can clearly still be consistent with the measured signal.

This visualization also illustrates in a very transparent way that a mere fit using the violins could never result in the spectra shown in yellow, implying that this analysis method cannot be sufficient when including cosmological constraints. As explained in section 5.3, the reason is that the violins no longer contain information about the likelihood for very small GWB spectra, where the signal is absorbed in the finely tuned pulsar-intrinsic red noise parameters $\{A_a, \gamma_a\}$.

⁷ The GWB posterior distributions are illustrated using the mean and standard deviation of 1000 spectra randomly drawn from their respective posterior distributions. The shaded areas hence correspond to the 1σ preferred spectral amplitudes.

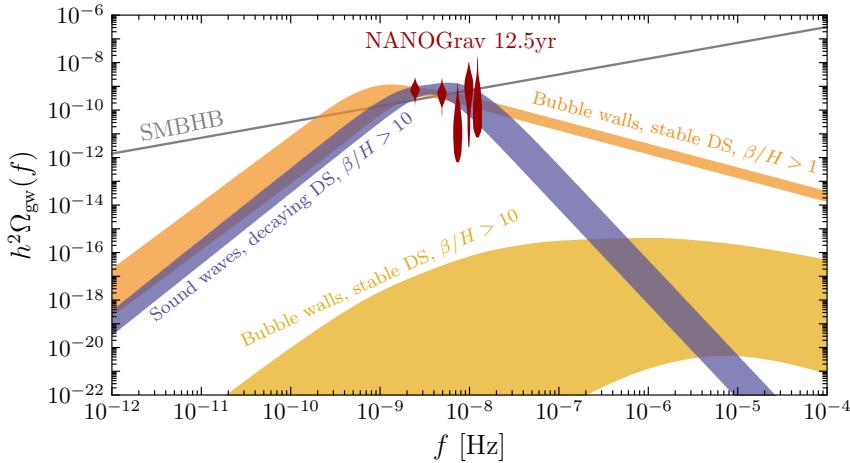


Figure 5.6: Envelopes of posterior distributions (at 1σ) of GWB spectra for parameter scans assuming bubble wall collision spectra from a stable DSPT with $\beta/H > 1$ (orange) and $\beta/H > 10$ (yellow), as well as for a decaying DS (violet spectra). The GW from SMBHBs with $A_{\text{SMBHB}} = 1.53 \times 10^{-15}$ [238] is depicted as a gray line for comparison.

5.C DETAILS ON THE CALCULATION OF BAYES FACTORS

5.C.1 The product space method

The Bayes factor between two models \mathcal{H}_1 and \mathcal{H}_0 is given by the evidence ratio

$$\mathcal{B}_{01} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}, \quad (5.23)$$

where \mathcal{Z}_i is the evidence of model i . The calculation of Bayes factors in this chapter relies on the product space method, which we briefly review here for completeness [246–249]. In order to compare \mathcal{H}_0 and \mathcal{H}_1 , a hyper-model \mathcal{H} is introduced whose parameter space is given by the Cartesian product $\boldsymbol{\theta}_{\mathcal{H}}$ of the two sub-models' parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ as well as an additional model index n that can formally run from -0.5 to $+1.5$. The key idea of this method is that the model index n can be treated as an additional, continuous parameter that is sampled over as any other parameter in an MCMC chain (while sampling over discrete parameters is technically more challenging). Whenever the hyper-model is evaluated, the underlying algorithm still simply casts the model index to either 0 or 1 , corresponding to one of the two sub-models. For a given model index n , the hyper-model parameter space is partitioned in an active part, which is used to evaluate the respective likelihood of the sub-model, and an inactive part. The posterior odds ratio \mathcal{P}_{01} is then the relative amount of chain entries in model 1 compared to model 0 , from which the Bayes factor \mathcal{B}_{01} between the two models can be deduced.

Extending the model parameter space with index n

To demonstrate in some more detail how this procedure can be used to calculate a Bayes factor, consider the posterior probability for the model index n

$$\begin{aligned} p(n|\text{data}, \mathcal{H}) &= \int d\boldsymbol{\theta}_{\mathcal{H}} p(\boldsymbol{\theta}_{\mathcal{H}}, n|\text{data}, \mathcal{H}) \\ &= \frac{1}{Z_{\mathcal{H}}} \int d\boldsymbol{\theta}_{\mathcal{H}} p(\text{data}|\boldsymbol{\theta}_{\mathcal{H}}, n, \mathcal{H}) p(\boldsymbol{\theta}_{\mathcal{H}}, n|\mathcal{H}). \end{aligned} \quad (5.24)$$

Here the first equality states that the posterior for n can be obtained by marginalizing the posterior for n and $\boldsymbol{\theta}_{\mathcal{H}}$ over the hyper-model parameters. The second equality follows from Bayes' theorem, where $Z_{\mathcal{H}}$ is the hyper-model evidence, the first term in the integrand is the likelihood and the second one is the prior for $\boldsymbol{\theta}_{\mathcal{H}}$ and n . The hyper-model evidence is unknown and difficult to obtain, but of no further importance, as we are interested in the posterior odds ratio \mathcal{P}_{01} between the two models. For a fixed n we can factorize

$$p(\boldsymbol{\theta}_{\mathcal{H}}, n|\mathcal{H}) = p(\boldsymbol{\theta}_n|\mathcal{H}_n) p(\boldsymbol{\theta}_{\bar{n}}|\mathcal{H}_{\bar{n}}) p(n|\mathcal{H}). \quad (5.25)$$

This factorization of the prior makes the aforementioned distinction between active (first factor) and inactive (second factor, where \bar{n} refers to all parameters not contained in model n) parameters explicit, which are not correlated with each other as the sub-models are distinct. The last factor is an overall, subjective prior for the respective sub-model. The two last factors do not depend on the active parameters $\boldsymbol{\theta}_n$. Inserting these expressions into the definition for the posterior odds ratio for the model index n , we get

$$\begin{aligned} \mathcal{P}_{01} &\equiv \frac{p(n=1|\text{data}, \mathcal{H})}{p(n=0|\text{data}, \mathcal{H})} \\ &= \frac{Z_{\mathcal{H}} \int d\boldsymbol{\theta}_{\mathcal{H}} p(\text{data}|\boldsymbol{\theta}_{\mathcal{H}}, n=1, \mathcal{H}) p(\boldsymbol{\theta}_{\mathcal{H}}, n=1|\mathcal{H})}{Z_{\mathcal{H}} \int d\boldsymbol{\theta}_{\mathcal{H}} p(\text{data}|\boldsymbol{\theta}_{\mathcal{H}}, n=0, \mathcal{H}) p(\boldsymbol{\theta}_{\mathcal{H}}, n=0|\mathcal{H})} \\ &= \frac{p(n=1|\mathcal{H})}{p(n=0|\mathcal{H})} \frac{\int d\boldsymbol{\theta}_1 p(\text{data}|\boldsymbol{\theta}_1, \mathcal{H}_1) p(\boldsymbol{\theta}_1|\mathcal{H}_1)}{\int d\boldsymbol{\theta}_0 p(\text{data}|\boldsymbol{\theta}_0, \mathcal{H}_0) p(\boldsymbol{\theta}_0|\mathcal{H}_0)} \frac{\int d\boldsymbol{\theta}_{\bar{1}} p(\boldsymbol{\theta}_{\bar{1}}|\mathcal{H}_{\bar{1}})}{\int d\boldsymbol{\theta}_{\bar{0}} p(\boldsymbol{\theta}_{\bar{0}}|\mathcal{H}_{\bar{0}})} \\ &= \underbrace{\frac{p(n=1|\mathcal{H})}{p(n=0|\mathcal{H})}}_{\equiv \Pi_{01}} \times \underbrace{\frac{Z_1}{Z_0}}_{\equiv \mathcal{B}_{01}}. \end{aligned} \quad (5.26)$$

In the last step we used that the inactive parameters, denoted by a bar, do not contribute to the sub-model evidence Z_n . A marginalization over their priors therefore gives one for both sub-models and the last factor in eq. (5.26) equals one. As \mathcal{P}_{01} is just the ratio of the number of chain entries after the burn-in period of sub-model 1 compared to sub-model 0, and the model weight ratio Π_{01} can be set as a model prior ratio when starting the MCMC chain, the Bayes factor is obtained by multiplying the posterior odds ratio with the inverse model weight ratio,

$$\mathcal{B}_{01} = \mathcal{P}_{01} \times \Pi_{01}^{-1}. \quad (5.27)$$

The posterior odds ratio for model 0 vs. model 1

The Bayes factor between models

5.C.2 Uncertainties of the computed Bayes factors

Our goal is to calculate the Bayes factor as accurately as possible. This can be achieved when the posterior odds ratio \mathcal{P}_{01} is close to one, i.e. if the hyper-model scan spends about the same amount of time in the two sub-chains. We therefore set the model weight ratio Π_{01} to the inverse of the expected Bayes factor and iterate over different weight ratios until the posterior odds ratio is close to one. In practice, this iterative procedure is still an intricate problem due to long runtimes. When the two models that are compared favor rather different regions of the available pulsar-intrinsic red noise parameter space, jumping from one sub-model to the other is initially unlikely. Hence, the burn-in period is long and uncertainties in the computed Bayes factors increase.

Most accurate results for $\Pi_{01} \rightarrow 1$

Even though our main aim is to calculate Bayes factors for different DSPT scenarios compared to the no-CURN hypothesis, it is faster and more reliable to first compare a given DSPT scenario with the SMBHB hypothesis as given in eq. (5.2). This speeds up the burn-in phase by an $\mathcal{O}(10)$ factor, and results in a more precise computation of the Bayes factor because the posterior distributions for the pulsar-intrinsic red noise parameters $\{A_a, \gamma_a\}$ are very similar between these two scenarios. As the Bayes factor between SMBHB and no-CURN hypotheses is known, $\log_{10} \mathcal{B}_{\text{SMBHB/no-CURN}} = 4.5(9)$ [238], and since Bayes factors are multiplicative, $\mathcal{B}_{02} = \mathcal{B}_{01} \times \mathcal{B}_{12}$, we can then simply rescale our results comparing to SMBHB to a Bayes factor that compares to the no-CURN hypothesis. Still, even with an informed choice of Π_{01} and using the method outlined above, the chains take several days before the Bayes factor converges. A publicly available implementation of the described pilot run in ref. [248] to speed up the calculation of Bayes factors would therefore be highly appreciated.

Speeding up the burn-in

Note, however, that the method of first comparing to the SMBHB hypothesis does not work if cosmological constraints do not allow a significant GWB in a DSPT model, like for a stable DS with a strong lower bound on β/H . Here, the DSPT models favor similar regions in pulsar-intrinsic red noise parameter space as in the case of the no-CURN hypothesis and a direct comparison would be more advantageous. The Bayes factors that we calculate in these cases therefore have larger uncertainties.

Larger uncertainties for weak signals

Computing Bayes factors inherently involves statistical uncertainties, particularly due to the finite length of the underlying chains. We ensured that this uncertainty is well-managed by calculating the Bayes factor as a function of the number of drawn samples. Using 5×10^6 samples from the hyper-model (including both sub-models) and conservatively discarding the first 25 % due to burn-in, the Bayes factors all converged to a relative uncertainty of a factor of up to ~ 2 . The convergence rate however depends sensitively on the precise value of the model weights Π_{01} as mentioned above. We therefore made sure that the number of

Statistical uncertainties are under control

samples of both models differed by no more than $\mathcal{O}(10\%)$, which requires to (iteratively) find the model weights Π_{01} up to one decimal place.

The continuous lines in fig. 5.4, depicting the Bayes factor expected prior dependencies, are further prone to uncertainties due to the reduction of points after adapting the prior, cf. appendix 5.C.4. This is particularly relevant when this reduction is large, i.e. when the Bayes factor is reduced by a significant amount, e.g. for a GWB from bubble wall collisions and a large lower boundary for β/H . Taken together, these uncertainties are the reason for the differences compared to the individually calculated Bayes factors for a given prior choice (solid dots).

5.C.3 Relating Bayes factors to p -values and Z -scores

Bayes factors can be expressed in terms of a “ Z -score” to describe the (im)probability of the null hypothesis [181, 281], more commonly known as the “number of sigmas” with which some measured quantity deviates from its expectation value. The probability of the null hypothesis can be interpreted as $p(0|\text{data}) = 1 - p(1|\text{data})$ in a frequentist’s manner, if one asserts that there is no other possible model to explain the data. If one further interprets the posterior odds ratio to be the ratio of these probabilities $\mathcal{P}_{01} = p(1|\text{data})/p(0|\text{data})$, one obtains $p(0|\text{data}) = 1/(1 + \mathcal{P}_{01})$. This can be interpreted as a p -value, i.e. the probability to measure data as extreme as the one observed if the null hypothesis were indeed correct. Assuming equal prior probabilities for the two models under comparison, i.e. $\Pi_{01} = 1$, the p -value reads $p = 1/(1 + \mathcal{B}_{01})$, which formally corresponds to $\mathcal{P}_{01} = \mathcal{B}_{01}$. We convert the obtained p -value to a Z -score using a one-tailed Gaussian,

$$Z = \Phi^{-1}(1 - p) = \Phi^{-1}\left(\frac{1}{1 + 1/\mathcal{B}_{01}}\right), \quad (5.28)$$

where Φ^{-1} is the inverse of the cumulative density function of the standard normal distribution with zero mean and unit standard deviation.

For Bayes factors \mathcal{B}_{01} below 1 we replace \mathcal{B}_{01} by its inverse, in order to express the tension between data sets if the null hypothesis is a better explanation for the observed data than a given, more complicated model. In that case the interpretation of Z is related to the probability of obtaining a signal as low as observed if the signal model were indeed correct.

5.C.4 Influence of the prior choice on the Bayes factor

The Bayes factor correction R

We now want to investigate the effect of a change in prior $\pi(\boldsymbol{\theta}) \rightarrow \tilde{\pi}(\boldsymbol{\theta})$ on the Bayes factor $\mathcal{B}_{01} \rightarrow \mathcal{B}_{0\tilde{1}}$ (in this section, we denote for simplicity the model parameters of model 1 by $\boldsymbol{\theta}$). Keeping the likelihood and its

normalization as well as the number of model parameters unchanged, the Bayes factors differs by a factor

$$R \equiv \frac{\mathcal{B}_{0\tilde{1}}}{\mathcal{B}_{01}} = \frac{\mathcal{Z}_{\tilde{1}}}{\mathcal{Z}_0} \cdot \frac{\mathcal{Z}_0}{\mathcal{Z}_1} = \frac{\int d\boldsymbol{\theta} \mathcal{L}(\boldsymbol{\theta}) \tilde{\pi}(\boldsymbol{\theta})}{\int d\boldsymbol{\theta} \mathcal{L}(\boldsymbol{\theta}) \pi(\boldsymbol{\theta})}. \quad (5.29)$$

Flat, proper priors (as assumed throughout this thesis) take a very simple form satisfying

$$1 = \int d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) = \frac{1}{V_\pi} \int_{V_\pi} d\boldsymbol{\theta} \quad \text{and} \quad 1 = \int d\boldsymbol{\theta} \tilde{\pi}(\boldsymbol{\theta}) = \frac{1}{V_{\tilde{\pi}}} \int_{V_{\tilde{\pi}}} d\boldsymbol{\theta}, \quad (5.30)$$

where V_π is the prior volume. In this case, the Bayes factor ratio simplifies to

$$R = \frac{V_\pi}{V_{\tilde{\pi}}} \frac{\int_{V_{\tilde{\pi}}} d\boldsymbol{\theta} \mathcal{L}(\boldsymbol{\theta})}{\int_{V_\pi} d\boldsymbol{\theta} \mathcal{L}(\boldsymbol{\theta})}. \quad (5.31)$$

The effect of changing the range of a flat prior can thus be understood intuitively: Increasing the prior into a region where the likelihood is negligible comes with the cost of increased prior volume, while the posterior integral is largely unaffected. If the likelihood were instead globally flat, an increase in prior volume would have no effect on the Bayes factor, as the increase in the posterior would just be compensated by the increased prior volume. This simply reflects the well-known feature of Bayesian statistics to disfavor unnecessary model complexity. In other words, the cost of introducing a new parameter depends on the coverage of the prior volume with the posterior. If the posterior of this parameter is flat, it can be introduced without changing the Bayes factor. If it however needs to be fine-tuned to fit the data, i.e. if its posterior is only a thin peak, the coverage of the prior volume is low, reducing the Bayes factor.

R compares the posterior-to-prior volume ratio change for two prior choices

The above considerations allow us to reduce the prior ranges after having computed a Bayes factor for some prior ranges that we initially set too wide, without the need to start a new MCMC chain. This is possible as the chain entries are distributed following the model likelihood $\mathcal{L}(\boldsymbol{\theta})$ (being proportional to the posterior for a flat prior). The ratio of integrals in eq. (5.31) thus reduces to a ratio of chain entries, thereby changing the Bayes factor as

$$R \rightarrow \frac{V_\pi}{V_{\tilde{\pi}}} \frac{N_{\tilde{\pi}}}{N_\pi}, \quad (5.32)$$

For flat priors, R can be computed easily

where $N_{\tilde{\pi}}$ and N_π are the number of chain entries enclosed within the prior volumes $V_{\tilde{\pi}}$ and V_π respectively. Note that this “a-posteriori” change of the priors is not more than a tool to quickly *estimate* the effect of slightly reducing of the prior volume. As soon as one cuts away a region of parameter space that was initially well-covered by the prior, the above approximation comes with an increased statistical error due to a potentially low number of samples $N_{\tilde{\pi}}$. This can for instance

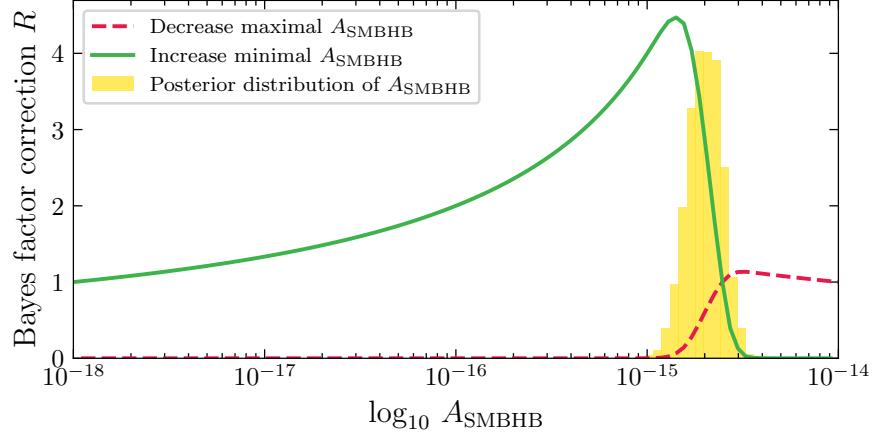


Figure 5.7: Bayes factor dependence on the A_{SMBHB} prior range for a SMBHB signal when increasing the lower boundary (green line) and lowering the upper boundary (red line). The posterior distribution of the SMBHB GWB spectral amplitude A_{SMBHB} is shown in yellow.

be seen by looking at the red line in fig. 5.4. It should also be noted that the uncertainty of this approximation does not scale with $1/\sqrt{N_{\tilde{\pi}}}$ due to the immanent properties of importance sampling which were required to evaluate the ratio of posterior integrals in the first place. A considerable reduction of the prior therefore comes with a significant error, in practice still requiring the computation of a full new MCMC chain with the updated prior ranges.

The effect of changing the prior range is also illustrated in fig. 5.7, where we show the Bayes factor ratio for SMBHBs with different choices of lower and upper prior boundaries on the amplitude A_{SMBHB} . As the posterior for the amplitude peaks between 10^{-15} and $10^{-14.5}$, the lower boundary of 10^{-18} on the log prior could be doubted to be a good choice, cf. ref. [216]. If one instead chose the lower boundary to lie at 10^{-15} , a factor four increase in the SMBHB's model evidence would be expected. Only lowering the upper prior boundary from 10^{-14} down to $10^{-14.5}$, on the other hand, would barely change the Bayes factor as this cuts away only a tiny amount of the prior parameter space that is favored by the posterior distribution.

Evidence for SMBHB can grow by a factor four

5.C.5 Influence of priors on the credibility of a DSPT

Interpreting the results of Bayesian statistics can be challenging given that posteriors and therefore also model evidences and Bayes factors are prior-dependent (even though this of course makes sense in the Bayesian framework, where prior beliefs should be updated when taking into account measured data). Not the least to assess the robustness of our main findings we therefore want to investigate the influence of prior choices on the Bayes factors. Concretely, let us get back to the

main question brought up in section 5.5.3: does the discrepancy between the Bayes factors for a decaying DSPT and the SMBHB interpretation, cf. fig. 5.4, have anything to do with the goodness-of-fit of the two explanations – or is it completely dominated by prior-volume effects?

In order to answer this question, we used a twofold approach: First we identified the region of parameter space that results in the highest posterior probabilities. Then we started two chains, one where all priors were constrained to only cover the best-fit regions⁸ and a second one in which we fixed all model parameters to their best-fit values⁹ except for α , for which we adopted a log prior in the range $[10^{-3}, 10^1]$. In doing so the prior in the second approach spans over four orders of magnitude, just as the prior for $A_{\text{SMBHB}} \in [10^{-18}, 10^{-14}]$, which allows for a comparison of the two models on a similar footing.

In the first analysis we obtained a Bayes factor of ~ 5 between the decaying DSPT and SMBHB interpretation in *favor* of the DSPT hypothesis. The second analysis found a Bayes factor of ~ 1.4 , again in (very slight) favor of the decaying DSPT interpretation. We checked explicitly that this matches the expected Bayes factors from an “a-posteriori” prior change as described in appendix 5.C.4. Both Bayes factors are also consistent with each other, since the one of the second analysis could be enhanced by a factor of 4 if one decreased the prior range to one decade, $\alpha \in [10^{-2}, 10^{-1}]$, confirming the ~ 5 Bayes factor of the first analysis. To make this a fair comparison it should however be noted that also the evidence of the SMBHB hypothesis can be boosted by a relative factor of 4 by decreasing the prior range of A from $[10^{-18}, 10^{-14}]$ to $[10^{-15}, 10^{-14}]$, cf. fig. 5.7 in appendix 5.C.4.

We therefore conclude that a GWB with a broken power-law spectrum with fixed slopes $\Omega_{\text{sw}}(f) \propto f^3$ and $\Omega_{\text{sw}}(f) \propto f^{-4}$ for low and high frequencies respectively, can in principle fit the common red spectrum slightly better than a featureless single power-law with spectral index $\gamma_{\text{SMBHB}} = 13/3$. As discussed further in section 5.5.3, the origin of the discrepancy between the highest Bayes factors depicted in fig. 5.4 and the reference value of $\sim 10^{4.5}$ is thus indeed due to prior volume effects.

*An equal footing
for model priors*

*DSPTs can fit the
data slightly better
than SMBHBs*

*Prior choices result
in the difference
between $\mathcal{B} = 10^{4.5}$
and $\mathcal{B} = \mathcal{O}(200)$*

⁸ We chose the following log prior ranges in this case: $\alpha \in [0.02, 0.04]$, $\beta/H \in [1, 3]$, $T_p/\text{MeV} \in [10, 100]$, $\xi_p \in [0.01, 0.1]$, $\tau_\phi/\text{s} \in [0.01, 0.1]$.

⁹ Specifically, we chose $\beta/H = 10^{0.3} = 2$, $T_p = 10^{-1.6} \text{ GeV} = 25 \text{ MeV}$, $\xi_p = 0.01$, $\tau = 10 \text{ ms}$. These values were chosen somewhat arbitrarily and are not the result of a precise maximization of the posterior probability. It is thus conceivable that a slightly better fit to the common red spectrum could be obtained by fine-tuning the parameter points.

5.C.6 Priors for the Bayesian model comparison

Parameter	Description		Prior	Comments
PSR-intrinsic red noise				
A_a	Red noise power-law amplitude	log Uniform $[-20, -11]$	one parameter per pulsar	
	Red noise power-law spectral index	Uniform $[0, 7]$	one parameter per pulsar	
Supernovae Black Hole Binaries (SMBHBs)				
ASMBHB	Red noise power-law amplitude	log-Uniform $[-18, -14]$	one parameter for PTA	
Stable dark sector phase transition, sound wave or bubble wall collision spectrum				
α	Phase transition strength	log-Uniform $[-5, 1]$	one parameter for PTA	
	Inverse timescale	log-Uniform $[0, 3]$, $[\log_{10} 3, 3]$, $[\log_{10} 5, 3]$, $[\log_{10} 7, 3]$, $[1, 3]$	one parameter for PTA	
T_p/GeV	SM temperature at percolation	log-Uniform $[-4, 1]$	one parameter for PTA	
	DS temperature ratio at percolation	log-Uniform $[-2, 1]$	one parameter for PTA	
	DS degrees of freedom	Constant 1	one parameter for PTA	
D	Dilution factor	Constant 1	one parameter for PTA	
			one parameter for PTA	
Decaying dark sector phase transition, sound wave spectrum				
α	Phase transition strength	log-Uniform $[-3, 1]$	one parameter for PTA	
	Inverse timescale	log-Uniform $[0, 3]$, $[\log_{10} 3, 3]$, $[\log_{10} 5, 3]$, $[\log_{10} 7, 3]$, $[1, 3]$	one parameter for PTA	
T_p/GeV	SM temperature at percolation	log-Uniform $[-4, 0]$	one parameter for PTA	
	DS temperature ratio at percolation	log-Uniform $[-3, 1]$	one parameter for PTA	
	Dark Higgs lifetime	log-Uniform $[-6, 2]$	one parameter for PTA	
ξ_p	DS degrees of freedom	Constant 1	one parameter for PTA	
	Dilution factor	Constant 1	one parameter for PTA	
	m_ϕ/MeV	Constant 5	one parameter for PTA	

Table 5.1: Table showing the model parameters together with their respective prior ranges.

6

HUNTING WIMPS WITH LISA

This chapter is based on the following publication:

- [2] T. Bringmann, T. E. Gonzalo, F. Kahlhoefer, J. Matuszak, and C. Tasillo, *Hunting WIMPs with LISA: correlating dark matter and gravitational wave signals*, *JCAP* **05** (2024) 065, [[2311.06346](#)]

*Snap back to reality
Ope, there goes gravity*

— Eminem

6.1 INTRODUCTION

Dark matter (DM) is known to be the dominant form of matter in the universe, but it has so far evaded any attempt of detection in the laboratory or by other non-gravitational means [39]. These null results have cast doubt on the so-called WIMP miracle, where DM is produced from the thermal bath of SM particles in the early universe, and which for a long time has been used to motivate sizable couplings between DM particles and the SM. Indeed, it has been shown that thermal freeze-out may happen entirely within an extended DS, such that the observed DM relic abundance $\Omega_{\text{DM}} h^2 \simeq 0.12$ [12] can be reproduced without the need for any sizable couplings between the dark and the visible sector [224]. These so-called secluded DM models pose a great challenge for laboratory searches due to their apparent lack of testable predictions.

At the same time, GW observatories have opened a completely new window into the universe, making it possible to observe objects and phenomena that affect visible matter only through gravity. The proposed LISA mission [282] will extend this window to the mHz frequency range, allowing in particular for the observation of a stochastic GWB that would be connected to a strong FOPT close to the electroweak scale [11, 27, 178], see fig. 2.8. LISA therefore raises new hopes to detect DSs that are otherwise unobservable.

Over the past few years, FOPTs in DSs have been studied in great detail [1, 4, 25, 26, 283], and various correlations between GW signals and the phenomenology of DM have been explored [186, 254, 255, 284–

*DM could have
frozen out straight
from a DS*

*LISA will probe
electroweak
temperatures*

There's a correlation between Ω_{DM} and f_{peak} since...

... $f_{\text{peak}} \propto v_\phi$...

[293]. The conclusion of these studies is that it is difficult to robustly predict the expected amplitude of the GW signal for a given DM model, because strong PTs often only happen in special regions of parameter space. In other words, it appears generally challenging to identify a strong correlation between the GW *amplitude* and the DM abundance. In this chapter, we instead focus on the *peak frequency* of the GW signal and show that it can be tightly correlated with the predicted DM relic abundance. Intriguingly, when imposing the observed value of $\Omega_{\text{DM}} h^2 = 0.12$ and focusing on GW signals strong enough to be potentially observable, we predict a GW peak frequency that falls right into the most sensitive range of LISA.

Before describing our analysis in detail, let us provide a rough sketch of the argument. We consider a DS comprised of a fermionic DM candidate χ charged under a new $U(1)'$ gauge group that is spontaneously broken by the vev v_ϕ of a new dark Higgs field. It is well known that strong PTs can occur in this model for a sufficiently large gauge coupling [294, 295]. All newly introduced particles are massless before symmetry breaking and acquire a mass proportional to v_ϕ afterwards. The dark gauge boson A' (a.k.a. dark photon) and the dark Higgs boson ϕ are generally unstable against decays into SM particles, but χ is stable and may obtain a sizable relic abundance through thermal freeze-out. If the spontaneous symmetry breaking occurs in a FOPT, bubbles of the new phase will nucleate spontaneously, expand and collide (cf. chapter 4). This process perturbs the dark plasma and leads to the emission of GWs, with a present-day peak frequency very roughly given by [178]

$$f_{\text{peak}} \simeq 10 \text{ mHz} \left(\frac{\beta/H}{100} \right) \left(\frac{T_p}{1 \text{ TeV}} \right). \quad (6.1)$$

Here β/H denotes the speed of the PT and T_p is the temperature of the SM heat bath at the time of percolation. For a not-too-strongly super-cooled DSPT, which is what we consider here, one expects $\beta/H \sim 100$ and $T_p \sim v_\phi$.

The relic density from thermal freeze-out, on the other hand, can in leading-order approximation be written as [38]

$$\Omega_{\text{DM}} \simeq 0.1 \frac{10^{-8} \text{ GeV}^{-2}}{\langle \sigma_{\text{ann}} v \rangle}, \quad (6.2)$$

with $\langle \sigma_{\text{ann}} v \rangle$ the thermally averaged DM annihilation cross section, cf. eq. (2.27). If the DM particles dominantly annihilate into the dark Higgs bosons ϕ , arising from the same dark Higgs field that generates the DM mass, it is parameterically of the form

$$\langle \sigma_{\text{ann}} v \rangle \sim \frac{y^4}{m_{\text{DM}}^2} \sim \frac{y^2}{v_\phi^2}, \quad (6.3)$$

where y denotes the DM Yukawa coupling. At first sight, this coupling is arbitrary, and hence the freeze-out mechanism does not predict a specific

DS mass scale. However, if we are interested in DSs that produce strong FOPTs and large GW signals, the dark gauge coupling g and the dark Higgs quartic coupling λ must be sizable, which implies that the dark Higgs boson mass m_ϕ cannot be much smaller than v_ϕ . At the same time, the observed DM relic abundance can only be obtained through DS freeze-out if the DM particle is not the lightest particle in the DS (or at least not much lighter than its annihilation products [66]). This, in turn, implies that y cannot be much smaller than unity, and hence $v_\phi \sim \text{TeV}$ once we require $\Omega_{\text{DM}} h^2 \sim 0.1$. Combining this with the conclusion from eq. (6.1), we thus expect a peak frequency of $f_{\text{peak}} \sim 10 \text{ mHz}$ —which, as advocated, lies right within the LISA band.

$$\dots \text{ and } \Omega_{\text{DM}} \propto v_\phi^2.$$

A possible concern with the simplified reasoning above is that a large Yukawa coupling will affect the effective potential and may possibly prevent a FOPT, or even destabilize the scalar potential [296]. We have also neglected the impact of additional DM annihilation channels involving dark photons. In our full analysis, we explore the entire parameter space of the model, calculating in detail the effective potential, the thermodynamic quantities characterizing the PT and the relic density from thermal freeze-out. We then identify viable combinations of the different DS couplings and show that the qualitative argument from above is confirmed by quantitative calculations. In order to further refine the analysis, we also perform parameter scans over all relevant model parameters—namely the three couplings g , λ and y and the dark Higgs vev v_ϕ , and we identify parameter points for which the correct DM abundance is obtained. Interpreting the sampling distributions for the model parameters as prior probabilities thus enables us to define “typical” model predictions and quantify the probability (in the Bayesian sense) of a detectable signal.

What could go wrong?

A significant focus of our analysis is to extend the simple argument sketched above to situations where the couplings are so weak that the dark and visible sectors do not necessarily share a common temperature, which would be maintained through (inverse) decays of SM and dark Higgs bosons. Indeed, even if the two sectors have the same temperature initially, the FOPT in the DS will change the temperature ratio, as the vacuum energy in the dark Higgs field is converted to rest mass and kinetic energy. This additional energy needs to be rapidly transferred to the SM in order to avoid a dilution of GW signals from late-time entropy injection [1, 4]. We calculate the dilution of the GWB and derive a lower bound on the portal coupling from the requirement that no significant dilution occurs. We show that the portal coupling required for this purpose is well below the sensitivity of laboratory experiments.

The thermalization of DS and SM

Finally, we explore what happens if the initial temperature ratio of the two sectors differs from unity. In this case the amplitude of the GW signal will change [25, 297, 298]—but the peak frequency remains almost unaffected, such that the estimate from above remains robust even for portal couplings that are too small to quickly (re-)thermalize the sectors after the transition. This conclusion is only modified if the

Outline of this chapter

portal coupling is so weak that the energy density of the DS cannot be depleted and starts to dominate the energy density of the universe.

The remainder of this chapter is structured as follows. In section 6.2 we introduce the model under consideration and identify the interesting regions of parameter space in which strong FOPTs can be predicted. In section 6.3 we calculate the DM relic density under the assumption that the dark and SM sector remain in thermal equilibrium throughout their evolution, and explore the correlation between the relic density and the GW signal. We revisit this assumption in section 6.4, and discuss in detail the processes that thermalize the DS with itself and with the SM. In section 6.5 we finally calculate the effect of inefficient thermalization on the GW signal. We consider the dilution due to entropy injection and show that for hot DSs a net enhancement of the GW amplitude can remain, while the peak frequency is essentially unaffected. We conclude in section 6.6 with a summary of our results and some remarks about their consequences. In two technical appendices, we provide details on the bubble wall velocity (appendix 6.A) and on the Boltzmann equations for entropy transfer (appendix 6.B).

6.2 DARK SECTOR MODEL

6.2.1 The effective potential

The model we study in this chapter was already introduced earlier in this thesis, in chapter 4.1.3, when we introduced the effective, temperature-dependent potential $V_{\text{eff}}(\phi, T_d)$. Starting with the tree-level Lagrangian in eq. (4.6) we found that the effective potential in eq. (4.22) allows for a spontaneous breaking of the imposed $U(1)'$ gauge symmetry, thus giving mass to the associated gauge boson A'_μ , the fermionic DM candidate χ and the dark Higgs boson ϕ . In chapter 4 we only provided a sketch of the effective potential for two representative cases resulting in a cross-over and a FOPT, in fig. 4.3. Now, as a first step of our analysis, we want to study the model parameter space in more detail.

For this purpose we use `TransitionListener` [4], an extension of the code `CosmoTransitions` [299], which takes care of the computation of the effective potential, the bounce action, as well as phase tracing and the calculation of the thermodynamic properties of the phase transition. For the moment, we will assume equal temperatures for the SM bath and the DS, i.e. we adopt a temperature ratio of

$$\xi \equiv \frac{T_d}{T} = 1. \quad (6.4)$$

The temperature that appears in the effective potential in eq. (4.22) and in the thermodynamic quantities discussed below can thus be identified

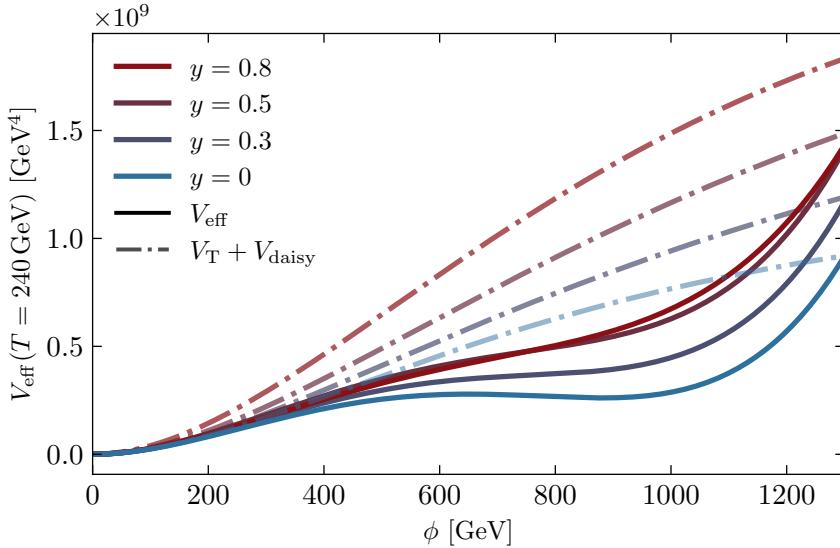


Figure 6.1: Total effective potential (solid) and finite temperature part (dot-dashed) for $g = 0.67$, $\lambda = 0.0035$, $v_\phi = 1 \text{ TeV}$ and $T = 240 \text{ GeV}$, for varying values of y .

with the temperature of SM photons. We discuss the general case with $\xi \neq 1$ in sections 6.4 and 6.5. The remaining $U(1)'$ parameter space hence depends on the three dimensionless couplings g , λ and y (i.e. the $U(1)'$ gauge coupling, the dark Higgs quartic coupling, and the DM fermion's Yukawa coupling) as well as the vev v_ϕ .

In previous works [4, 25], the $U(1)'$ model was already studied in great detail under the assumption that $y = 0$, i.e. assuming that there is no DM fermion present in the dark sector. Let us therefore focus on the impact of the Yukawa coupling y on the effective potential: In fig. 6.1 a plot of the effective potential for a choice of parameters $g = 0.67$, $\lambda = 0.0035$, $v_\phi = 1 \text{ TeV}$ and varying values of y can be found. We find that also in the presence of a fermion coupled to the dark Higgs field, the effective potential can still provide a barrier necessary for the occurrence of a FOPT. The larger the Yukawa coupling y , the more the development of the broken phase is delayed, however, due to the thermal corrections $V_T + V_{\text{daisy}}$ within the effective potential.

In addition to encoding the properties of the phase structure, the effective potential also provides information about the stability of the true vacuum after the PT occurs. In fact, a new feature becoming important for non-zero Yukawa couplings is that for low values of λ and g the potential can become unbounded from below [296]. To ensure vacuum stability we require that no deeper vacua are present at zero temperature. The requirement of a DSPT already implies that $V_{\text{eff}}(0) > V_{\text{eff}}(v_\phi)$. Hence, it is sufficient to check whether there exist vacua with lower potential energy for large field values, i.e. whether $V_{\text{eff}}(\phi) < V_{\text{eff}}(v_\phi)$ for $\phi \gg v_\phi$. In our analysis, we explicitly exclude such parameter points.

Let's start with the $\xi = 1$ case

Fermions destabilize the vacuum

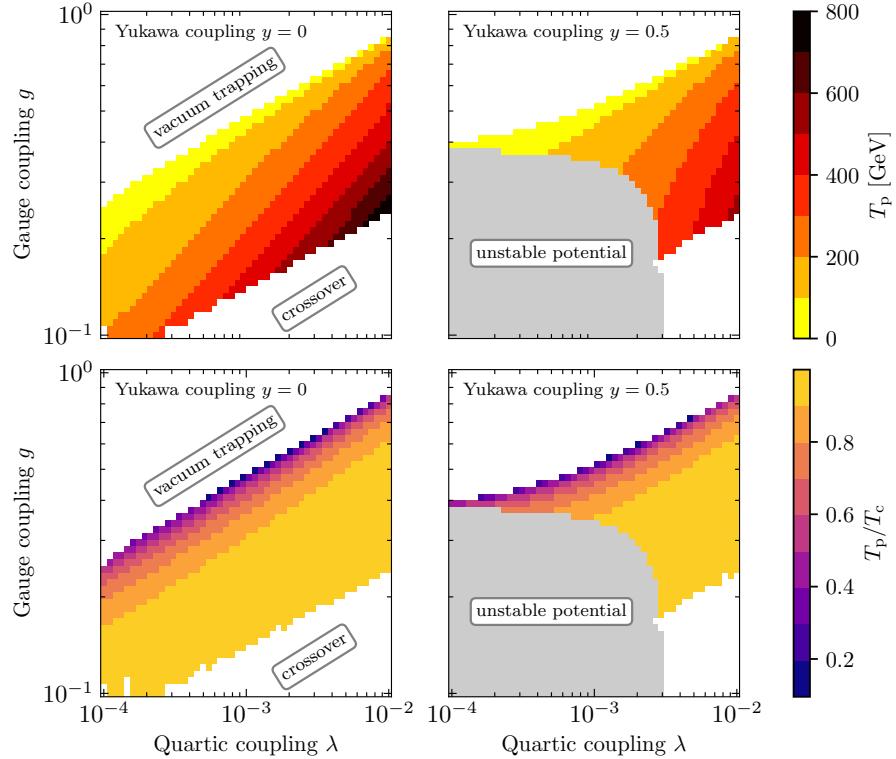


Figure 6.2: The percolation temperature T_p (top) and the ratio of percolation temperature and critical temperature T_p/T_c (bottom) in the $\lambda - g$ plane for Yukawa couplings of $y = 0.0$ (left) and $y = 0.5$ (right) and a vev of $v_\phi = 1 \text{ TeV}$. The colored band shows the parameter region where a first-order PT is possible.

It is well known that the one-loop, daisy-resummed calculation of the effective potential can suffer from large theoretical uncertainties, foremost sourced by a large renormalization scale-dependence [300]. A possibility to improve upon those uncertainties is to systematically resum higher orders of the thermal masses in the effective field theory framework of dimensional reduction [301]. In order to validate our simpler approach, we therefore also implemented our model in **DRalgo** [171], which automates the task of dimensional reduction. We calculate the critical temperature in both our four-dimensional implementation and the reduced three-dimensional theory for the parameter space where we expect a FOPT. In the regime where the effective field theory is valid ($T \gg m_\phi$) we find that the two results agree very well. We therefore conclude that we can use the computationally more economical approach of using the 1-loop, daisy-resummed effective potential stated in eq. (4.22) within the following calculations.

*DRalgo validated
 $V_{\text{eff}}(\phi)$*

6.2.2 Properties of the phase transition

Next, we want to study the dependence of the percolation temperature T_p (see eq. (4.28)) and the critical temperature T_c on the model parameters. We define the critical temperature as the temperature at which the minimum of the effective potential with non-vanishing vev becomes a global minimum.

The dependence of the percolation temperature T_p on the model parameters is shown in fig. 6.2. In the top panels T_p is displayed as a function of the quartic coupling λ and gauge coupling g for two values of the Yukawa coupling, $y = 0$ (left) and $y = 0.5$ (right). There is a strong correlation between the values of λ and g that produce a FOPT, with lower values of T_p in the top-left end of the allowed band, and higher values of T_p in the bottom-right. The disallowed areas correspond to parameter regions where the transition is not first-order or does not occur at all. These effects are better illustrated in the bottom panels, where the color scale indicates the ratio T_p/T_c in the same parameter plane. The amount of supercooling of the transition is largest when T_p is much lower than T_c and smallest when both temperatures almost coincide. For the points above the colored contours, the potential barrier becomes so large that the bubble nucleation rate is too low for the transition to reach percolation; the region below instead indicates a smooth crossover transition in which no bubbles form since the potential does not develop a barrier between the phases. For non-zero values of the Yukawa coupling y , the enhanced thermal corrections in the effective potential cause a delay of the development of the true vacuum (cf. fig. 6.1), thereby decreasing the value of T_p . The vacuum also becomes deeper due to the Yukawa coupling, which increases the tunneling rate close to the supercooled region, and thus slightly larger values of g are within the allowed band. The gray shaded regions, finally, indicate parameter combinations where the potential is unstable.

Supercooling for high g and low λ

In fig. 6.3 we show how the transition strength α (see eq. (4.29)) and transition speed β/H (see eq. (4.31)) depend on the model parameters, λ , g and y . The PT is relatively strong for most of the allowed region $\alpha \in (10^{-2}, 10^2)$ and it is particularly strong close to the supercooled limit, where percolation is delayed ($T_p \ll T_c$). On the other hand, the speed of the PT β/H becomes smaller in the supercooling limit, reaching values of $\beta/H \approx 10^2 - 10^3$. Both are indicators for a strong GW background.

Strong and slow transitions are possible

6.2.3 The gravitational wave spectrum

The spectrum of GWs in our scenario is produced dominantly through bulk fluid motion in the reheated plasma due to the large velocity-dependent friction from the emission of soft dark photons in the bubble wall, yielding a terminal bubble wall velocity [302–304]. A discussion of

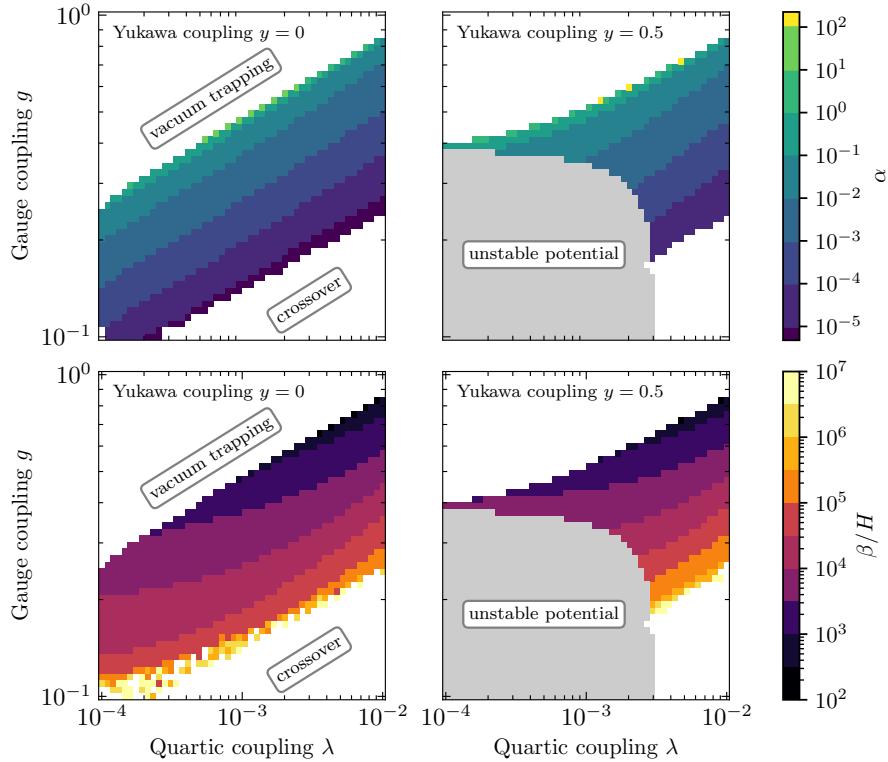


Figure 6.3: The transition strength α (top) and speed β/H (bottom) of the PT, in the $\lambda - g$ plane for $y = 0$ (left) and $y = 0.5$ (right), with $v = 1 \text{ TeV}$ and $\xi = 1$.

this argument can be found in appendix 6.A. As the case of runaway bubbles can hence be excluded, we neglect the contribution of bubble collisions to the GW signal. Since the onset of turbulence as a GW source is not yet understood well enough to make quantitative statements [178], and often requires complicated lattice simulations [191], we conservatively consider sound waves the only relevant source of GWs emitted during the PT. Therefore, the GW spectrum is exclusively determined by the previously discussed set of parameters $\{\alpha, \beta/H, T_p\}$. We use the semi-analytical approximation provided in eq. (4.33) to compute the peak frequency and GW spectrum from sound shell collisions.

A strong GW signal typically requires sizable α and values of β/H that are not too large. From the discussion of fig. 6.3, a strong FOPT implies large but perturbative values of both g and λ . Too small values of these two couplings would imply very large values of β/H and correspondingly weak GW signals, and even cause issues of vacuum stability (for large y , cf. right panel of fig. 6.3). This in turn induces an upper limit on the value of y as large values would cause an unstable vacuum for any perturbative value of g and λ . As will be seen below in section 6.3, successfully producing the right DM relic density requires $m_\chi \gtrsim m_\phi$, which implies a lower limit $y > 2\sqrt{\lambda}$. Lastly, the vev v_ϕ is chosen in a range that produces GWs in the frequency range of near-future GW observatories, such as LISA.

*Obtaining strong
GWs requires
strong couplings*

Consequently, we randomly draw parameters from distributions that are logarithmically flat within the following ranges: $0.1 \leq g \leq 1$, $10^{-4} \leq \lambda \leq 10^{-2}$, $0.01 \leq y \leq 0.7$ and $10^{-3} \text{ GeV} \leq v_\phi \leq 10^3 \text{ GeV}$. We then discard parameters that cause the vacuum to be unstable, that do not predict a FOPT, or for which the PT is too supercooled and never percolates, thereby removing 82% of the points drawn. The remaining 18% of parameter points all feature a FOPT with a corresponding GW signal. However, since the percolation temperature is very sensitive to small changes in the couplings, the PT is only strong enough to give an observable GW signal in certain small regions of parameter space. Indeed, only about 1% of parameter points from the original sample feature strong supercooling ($T_p/T_c < 0.5$).

*Strong supercooling
for 1% of points*

We can quantify the fine-tuning required to obtain an observable GW signal by interpreting our parameter scan as a sample drawn from the prior distributions of the parameters. We then find that out of the parameter points that give a FOPT, only about 0.8% would be observable with LISA, whereas this number increases to 10% if we select parameter points that give a strongly supercooled PT. For the parameter ranges that we consider (in particular of v_ϕ) none would be observable with PTAs or the Einstein Telescope. We note that these numbers do not correspond to rigorously calculated posterior probabilities, but rather rough estimates based on sampling densities. More precise estimates would require a different sampling strategy (see e.g. [305]), which is beyond the scope of this chapter.

*Out of these, 10%
can be seen with
LISA*

We emphasize that these numbers are largely independent of the choice of priors as long as we select only parameter points that predict any kind of FOPT. The probability to find parameter points that give a FOPT does however depend sensitively on the choice of priors. If we were to extend the prior ranges for all parameters to lower couplings, the volume of parameter space without a FOPT would grow significantly. Choosing for example $g > 0.01$ (instead of 0.1), $\lambda > 10^{-5}$ (instead of 10^{-4}) and $y > 10^{-3}$ (instead of 0.01) would decrease the fraction of parameter points with a FOPT from 18% to 6%. Out of these, 0.7% would be observable by LISA, which increases to 8.5% when considering only points with a strongly supercooled PT. As expected, our results are not very sensitive to different prior choices as we find that points that already have a FOPT have a roughly equivalent probability of being visible at LISA regardless of the parameter ranges. In later sections, we will discuss how these numbers change when imposing additional constraints on the DS, such as the relic density requirement.

*Dependence on
prior choices*

Finally, when studying the effects of thermalization in our model in section 6.4.2 it will be convenient to identify a benchmark scenario with the right properties for the PT and DM relic abundance. For reference the benchmark point is given in table 6.1.

g	λ	y	v_ϕ	m_χ	m_ϕ	$m_{A'}$
0.67	0.0035	0.62	430 GeV	189 GeV	36 GeV	288 GeV

Table 6.1: Benchmark point used for discussing the thermalization of visible and dark sector.

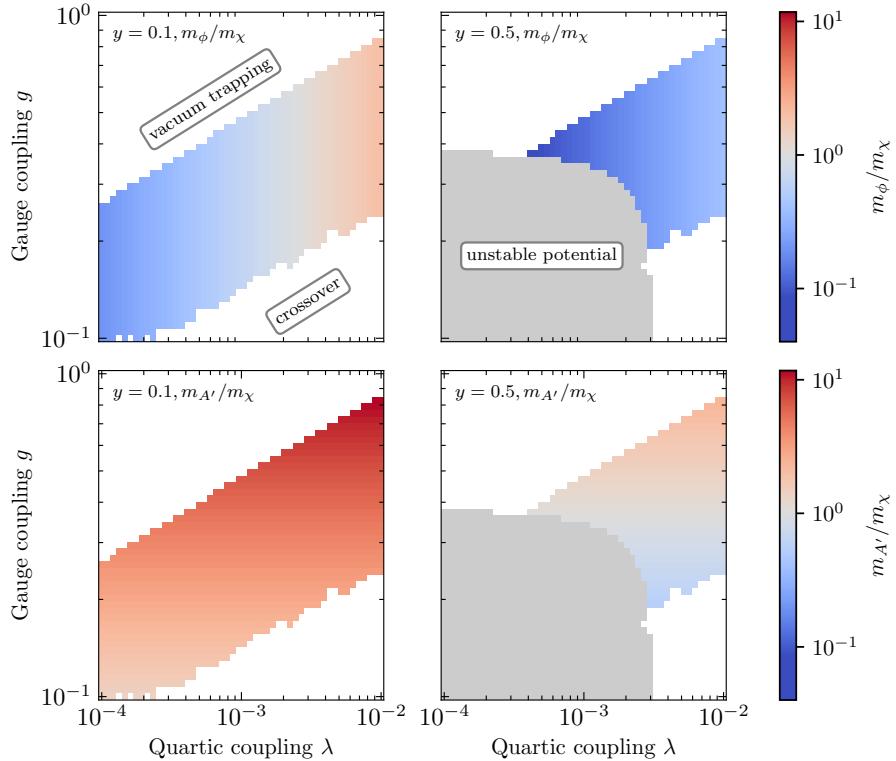


Figure 6.4: The upper (lower) panels show the ratio of the dark Higgs boson mass m_ϕ (the dark photon mass $m_{A'}$) to the mass of the DM fermion m_χ for $y = 0.1$ (left) and $y = 0.5$ (right), as a functions of the gauge coupling g and the self-interaction λ . Note that these ratios are independent of the dark Higgs vev.

6.3 THE DARK SECTOR RELIC DENSITY

During the PT, the DS particles χ , ϕ and A' all obtain masses proportional to the dark Higgs vev v_ϕ . In the parameter regions of interest for a strong FOPT, we generally find $g > \sqrt{2\lambda}$ and $g > y/\sqrt{2}$ and hence the dark photon is usually the heaviest state in the DS, cf. eq. (4.8b). Depending on the value of the Yukawa coupling y , the lightest DS particle will instead be either the DM fermion or the dark Higgs boson, as shown in fig. 6.4. The DS equilibrates soon after the PT (see section 6.4 for a more detailed discussion). Typically, the heaviest particles will then first drop out of equilibrium as their number densities become strongly suppressed. The relic abundance of the dark fermions χ is thus determined through a freeze-out process [306] in the usual way. We assume that the dark photon is unstable, decaying for example through

$$\begin{aligned} m_\chi < m_\phi < m_{A'} \\ \text{or} \\ m_\phi < m_\chi < m_{A'} \end{aligned}$$

kinetic mixing, and therefore does not contribute to the DM relic density (unlike the case studied in ref. [290]).

In our model there are three possible DM annihilation processes that are relevant for setting the DM abundance: $\chi\chi \rightarrow \phi\phi$, $\chi\chi \rightarrow \phi A'$ and $\chi\chi \rightarrow A'A'$. If the DM fermion is the lightest particle in the DS, annihilation into other DS states is kinematically forbidden for vanishing kinetic energy, such that the annihilation cross section becomes exponentially suppressed at low temperatures. In this so-called ‘forbidden’ regime [66], a relic abundance in accordance with observations requires that all mass scales must be correspondingly smaller, or the relevant couplings (much) larger. For the parameter values we are interested in here, it is therefore typically necessary for the DM particle to be heavier than the dark Higgs boson, which in turn requires a sizable Yukawa coupling y . For even heavier DM, with $2m_\chi \gtrsim m_\phi + m_{A'}$, the annihilation channel $\chi\chi \rightarrow \phi A'$ opens up. This process is a highly relevant contribution, once kinematically accessible, as it proceeds via an s -wave; the annihilation into a pair of dark Higgs bosons, $\chi\chi \rightarrow \phi\phi$, on the other hand, only proceeds via a p -wave.

Available DM annihilation channels

To compute the DM relic density, we have calculated the amplitudes for all three processes, see appendix 6.B.3, and implemented them in **DarkSUSY** [307], which calculates the thermal averages and solves the full Boltzmann equation [308]. While **DarkSUSY** allows precision calculations of the relic density in a fully secluded DS with a varying temperature ratio ξ between the dark and the SM sector, cf. ref. [279], we will set $\xi = 1$ for the purpose of this section. We will revisit this assumption of thermal equilibrium between the two sectors in section 6.4.

We solve Boltzmann equations with DarkSUSY

We show the results from the parameter scan described in section 6.2.3 in fig. 6.5. The three two-dimensional scatter plots show the correlation between the DM relic density $\Omega_{\text{DM}} h^2$, the peak frequency f_{peak} as well as the peak amplitude $\Omega_{\text{GW}}^{\text{peak}} h^2$. One can immediately see that Ω_{DM} and $\Omega_{\text{GW}}^{\text{peak}}$ are not tightly correlated (with a correlation coefficient of 0.20), while there exists a clear connection between the DM relic density and the peak frequency (with a correlation coefficient of 0.85). We can trace this correlation back to the fact that both quantities are determined by the dark Higgs vev v_ϕ (indicated by the color of each point). A smaller value of v_ϕ implies a smaller DM mass and therefore a larger annihilation cross section, which in turn results in a smaller relic density. At the same time, a smaller v_ϕ also implies a smaller percolation temperature, and hence a smaller peak frequency. The strength of the PT, on the other hand, depends on the details of the effective potential, and can vary over many orders of magnitude for any given value of v_ϕ .

Tight correlation between Ω_{DM} and f_{peak}

We complement these scatter plots by showing distributions of the derived quantities, in the form of histograms based on our random scan described above. For example, one can infer that most samples drawn in our setup correspond to a peak GW signal strength of $\Omega_{\text{GW}}^{\text{peak}} h^2 \approx 10^{-16}$, i.e. a few orders of magnitude below the PLI sensitivity of near-future

Typically $\Omega_{\text{GW}}^{\text{peak}} h^2 \approx 10^{-16}$

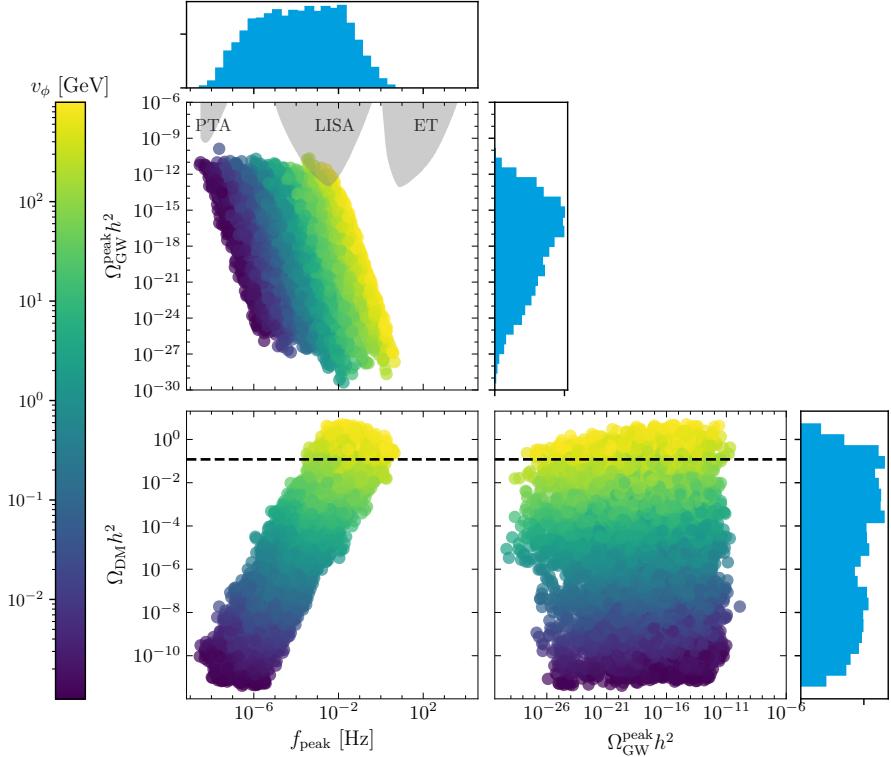


Figure 6.5: Scatter plots and 1D distributions of the DM density $\Omega_{\text{DM}}h^2$, the GW peak amplitude $\Omega_{\text{GW}}h^2$, and the peak frequency f_{peak} . For comparison, the dashed line shows the observed DM density, $\Omega_{\text{DM}}h^2 = 0.12$ [12]; gray shaded areas show the PLI sensitivities [25] of PTAs, LISA and ET, respectively.

GW observatories (indicated as gray shaded areas). Note also that the DM density caps at $\Omega_{\text{DM}}h^2 \approx 10$, which would already correspond to an overclosed universe; even higher values are avoided by our prior choice, in particular the upper bound on v_ϕ and the lower bound on y .

In fig. 6.6 we show the result of sharpening the relic density requirement by requiring that $0.06 \leq \Omega_{\text{DM}}h^2 \leq 0.12$. Demanding in this way that the fermionic DM candidate in our model constitutes the dominant form of DM, the predicted range of peak frequencies of the GW signal shrinks significantly—as expected from the discussion above. Interestingly, almost all viable parameter points now predict a peak frequency between 0.1 mHz and 100 mHz, largely overlapping with the frequency range to which LISA is sensitive. In fact, the peak frequencies for those parameter points that result in the strongest signal are the same as those where LISA is most sensitive. This striking correlation is a non-trivial feature of our model and constitutes one of our main results. Let us note that a few points remain that predict peak frequencies outside the LISA band. Much smaller values of f_{peak} , in particular, correspond to parameter points in the ‘forbidden’ regime, $m_\chi < m_\phi$, where DM annihilations are exponentially suppressed at small temperatures. Smaller values of v_ϕ (and hence smaller temperatures of the PT) can then still result in

Require
 $\Omega_{\text{DM}}h^2 \simeq 0.12$:
GWs in LISA
band!

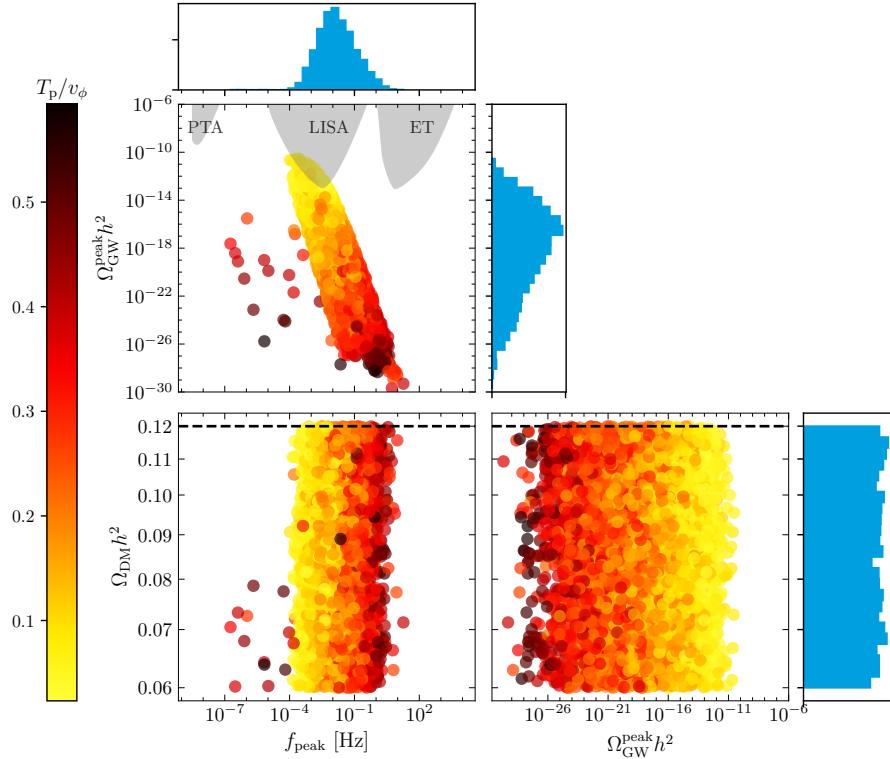


Figure 6.6: Scatter plots and 1D distributions as in fig. 6.5, but with the additional constraint $0.06 \leq \Omega_{\text{DM}} h^2 \leq 0.12$. Here, the color scale does not encode the vev, but the ratio of percolation temperature T_p to the vev v_ϕ , thus indicating the amount of supercooling.

the correct DM relic abundance, but only at the cost of a significant tuning between the various couplings (as reflected by the rareness of such parameter points, cf. the f_{peak} histogram at the top of the plot).

In contrast to fig. 6.5, where the color coding of each point represents the dark Higgs vev, the points in fig. 6.6 are colored according to the percolation temperature of the PT, normalized to the vev v_ϕ . Doing so allows us to confirm that the peak amplitude of the GW spectrum is determined primarily by the amount of supercooling. In other words, if the PT is delayed by a large potential barrier, the strength of the PT increases, yielding strong GW amplitudes (as expected from figs. 6.2 and 6.3). As discussed in section 6.2.3, the predictions for the PT properties vary a lot with small changes of the model parameters, and thus only certain regions of the parameter space predict a strong FOPT. For this reason, our model cannot in general guarantee a strong PT, and thus a GW signal that is visible with next-generation GW observatories.

Observability requires strong supercooling

We can make this statement more precise if we interpret the sampling distributions of the model parameters as prior probabilities (as we did in section 6.2.3), such that the density of points in figs. 6.5 and 6.6 can be interpreted as probability distributions for the observables under consideration. As before, this makes it possible to quantify the amount of fine-tuning required to obtain a strong FOPT, through the fraction

3% of points with $\Omega_{\text{DM}}h^2 \simeq 0.12$ can be seen with LISA

of points with a FOPT that predict a signal observable with LISA. If we do not impose the relic density requirement (fig. 6.5), only 0.8% of points with a FOPT predict a GW signal visible at LISA, whereas this fraction increases to 3% once the relic density requirement is included (fig. 6.6). If we restrict ourselves to parameter points with a strongly supercooled PT, the fraction of observable parameter points increases from 10% to 35%. Again, we have checked that this number is not very sensitive to our choice of parameter ranges.

6.4 THERMALIZATION OF THE TWO SECTORS

In this section we revisit the assumption that the temperature ratio of the dark and visible sectors is $\xi = 1$ throughout the PT. To do so, we first need to understand the evolution of the DS temperature during the PT, and convince ourselves that the DS quickly thermalizes with itself afterwards, such that the DS states remain in kinetic equilibrium with each other until after DS freeze-out (i.e. chemical decoupling). However, it is not necessarily the case that also the SM states are in kinetic equilibrium with the DS, such that their temperature may differ from the one of the DS both before and after the PT. We therefore discuss the various processes that allow for the exchange of energy and entropy between the dark and the SM sector, and the resulting Boltzmann equations. This enables us to identify the necessary portal couplings for efficient thermalization. For the case of delayed thermalization, after the end of the PT, we calculate the resulting dilution of the GW signal due to the injection of entropy into the SM thermal bath.

For the purpose of illustration, we will in this section consider a specific benchmark point that we selected from the random parameter scan discussed previously (see table 6.1). For $\xi = 1$, the parameters of this point lead to $\alpha = 0.258$, $\beta/H = 874$, $T_n = 39.7 \text{ GeV}$, $T_p = 39.1 \text{ GeV}$, $f_{\text{peak}} = 3 \text{ mHz}$, $\Omega_{\text{DM}}h^2 = 0.117$, and $\Omega_{\text{GW}}^{\text{peak}}h^2 = 3 \cdot 10^{-13}$. The rationale behind choosing this benchmark point is that (i) the observed DM relic abundance is reproduced (for $\xi = 1$), and that (ii) the PT is sufficiently strong in order to obtain an observable signal in LISA. We have explicitly checked that our choice is representative in the sense that other points fulfilling these two criteria lead to a very similar temperature evolution and resulting predictions.

A benchmark point

Let's drop the $\xi = 1$ assumption

6.4.1 The dark sector temperature

As the bubbles of the broken phase expand, more and more DS particles will pass through the bubble walls and enter the new phase. In the process, not only their rest masses but also their kinetic energies increase dramatically, by converting the vacuum energy of the dark Higgs field

stored in the false vacuum. Here we neglect the small fraction of the energy density that is converted into GWs and assume that the bubble walls have already reached their terminal velocity, such that no energy is needed for their acceleration. As we have learned in the previous section, in particular, the energy density of GWs produced in the PT is bounded by $\Omega_{\text{GW}}^{\text{peak}} h^2 < 10^{-10}$ and can therefore safely be ignored. We also neglect the effect of bubble filtering [287, 309], i.e. we assume that all DS particles can enter the new phase. This is a good approximation for sufficiently fast bubble walls, see appendix 6.A for details.

*We can ignore
bubble filtering*

Since the different particle species in the DS were all relativistic before the PT, their number densities immediately after the PT will be comparable, even though their masses will now be very different. Indeed, for strongly supercooled PTs the dark photons (and possibly also the DM particles) will typically have a large mass compared to the temperature of the plasma, such that their equilibrium number density would be Boltzmann-suppressed. In other words, right after the PT the DS finds itself far away from thermal equilibrium. Nevertheless, interactions between the different DS particles are rather strong, and hence the heavier particles are expected to annihilate rapidly into lighter ones, thereby restoring equilibrium.

*Particles behind
the wall are out of
equilibrium*

As we will show below, the time required to reach equilibrium is negligible compared to the duration of the PT, such that we can to a very good approximation define a DS temperature of the broken phase $T_{\text{d},\text{br}}$ immediately after the PT. This temperature is obtained from the temperature of the symmetric DS phase $T_{\text{d},\text{sym}}$ using energy conservation:

$$\rho_{\text{vac}}(T_{\text{d},\text{br}}) + \rho_{\text{DS}}(T_{\text{d},\text{br}}) = \rho_{\text{vac}}(T_{\text{d},\text{sym}}) + \rho_{\text{DS}}(T_{\text{d},\text{sym}}). \quad (6.5)$$

*Energy
conservation fixes
the broken-phase
temperature*

Here, $T_{\text{d},\text{br}}$ denotes the temperature in the broken phase and

$$\rho_{\text{DS}}(T_{\text{d},\text{br}}) = \frac{\pi^2}{30} g_{\text{DS}}(T_{\text{d},\text{br}}) T_{\text{d},\text{br}}^4, \quad (6.6)$$

where $g_{\text{DS}}(T)$ takes into account the T -dependence stemming both from thermal (field-dependent) masses and the minimum of the effective potential. Eq. (6.5) can easily be solved numerically for $T_{\text{d},\text{br}}$ for a given $T_{\text{d},\text{sym}}$.

In practice, we find that slightly different temperatures $T_{\text{d},\text{br}}$ of the broken phase are obtained when solving the equation taking $T_{\text{d},\text{sym}} = T_p$ or $T_{\text{d},\text{sym}} = T_n$. This is because the energy density of the broken phase redshifts differently from the symmetric phase. We have therefore implemented a more detailed calculation, which tracks the temperature of the symmetric and broken phases from bubble nucleation to percolation and applies eq. (6.5) at each time step to the fraction of the universe entering the broken phase. Here the energy in the bubble walls, which for relativistic bubble wall velocities redshifts like radiation [35], is included in the energy of the symmetric phase. We find that this more careful treatment gives very similar results to simply applying eq. (6.5) at

*We set $T_{\text{d},\text{sym}}$ to
percolation
temperature*

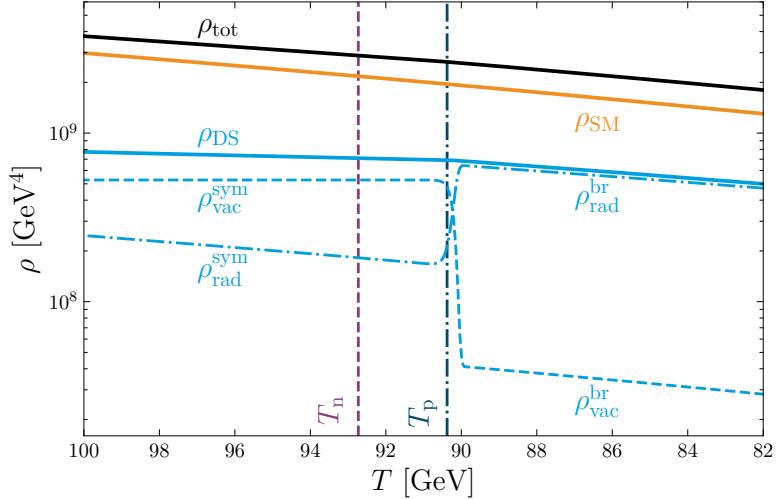


Figure 6.7: Contributions to the energy density around the PT for our benchmark point, as a function of the SM temperature and for a temperature ratio $\xi = 1$. The energy densities in the symmetric and broken phase of the DS have two contributions, namely the energy of the particles ('rad') and the potential energy of the scalar field ('vac').

$T_{d,sym} = T_p$. We therefore use the latter prescription in the following when computing the temperature of the DS after the PT.

For our benchmark point, we find that the energy density of the DS before the PT is dominated by vacuum energy (see fig. 6.7). In the broken phase, on the other hand, the vacuum energy is very small and quickly relaxes to a value very close to its zero temperature value as the temperature decreases further. This difference in vacuum energy leads to a substantial reheating of the DS, which, as a result, is hotter than the SM sector after the PT. For our benchmark point, we find that if the two sectors have equal temperature before the PT, in the broken phase the DS temperature will be larger by a factor of about 1.3. This reheating of the DS typically ensures that the dark Higgs bosons will be relativistic immediately after the PT.

Reheating of the DS leaves ϕ relativistic

6.4.2 Thermalization within the dark sector

In the discussion above we have assumed that the DS can be characterized by a common temperature shortly after the PT. To justify this approach, we need an estimate of the time τ required to reach this equilibrium state and show that it is sufficiently small. For this purpose, we calculate the interaction rate for each DS state X in thermal equilibrium:

$$\Gamma_X = \sum_Y \langle \sigma_{XY} v \rangle n_Y^{\text{eq}}, \quad (6.7)$$

where the sum is over all DS states Y , σ_{XY} denotes the total interaction cross section of X and Y , brackets denote thermal averaging (for simplicity calculated by assuming Boltzmann distributions) and n_Y^{eq} denotes the equilibrium number density of Y .

A total of 20 different processes contribute to the thermalization of the DS, the relative importance of which depends on the specific choice of parameters and the DS temperature. In the interest of brevity we refrain from stating the thermalization rates explicitly. Broadly speaking, we find that Γ_X is only a few orders of magnitude smaller than m_X . For example, dark Higgs bosons can thermalize via self-scattering, i.e. $\phi\phi \leftrightarrow \phi\phi$, for which the scattering cross section is $9\lambda^2/(8\pi s)$. For temperatures comparable to the dark Higgs boson mass, we have $s \approx 4m_\phi^2$ and $n_\phi \approx \zeta(3)m_\phi^3/\pi^2$, such that $\Gamma_\phi \sim 10^{-7}m_\phi$ for the benchmark point. Interactions of the dark Higgs bosons with dark fermions or dark photons benefit from the larger couplings $y, g \gg \lambda$, but suffer from a Boltzmann suppression if $T_p < m_\chi, m_{A'}$.

A rough estimate of the thermalization timescale is then obtained via

$$\tau = \max_X \Gamma_X^{-1}. \quad (6.8)$$

Given the timescale τ we can estimate the out-of-equilibrium fraction $F(t)$ of the universe, which describes the relative volume of the universe that entered the broken phase so recently that it has not had enough time to reach thermal equilibrium. The related false vacuum fraction $P(t)$, cf. eq. (4.28), describes the fraction of the universe which has not yet transitioned to the new phase, such that the true vacuum fraction is given by $P_t(t) = 1 - P(t)$. Its time derivative \dot{P}_t describes the rate with which the volume is transitioning to the true minimum of the potential for a given time t . We introduce the quantity

$$F(t) \equiv P(t - \tau) - P(t) > 0, \quad (6.9)$$

which can hence be interpreted as the volume fraction $\dot{P}_t \Delta t$ that just transitioned to the broken phase within the small thermalization time scale $\Delta t = \tau$, cf. fig. 6.7. The volume fraction F becomes small exactly when the thermalization timescale τ is small compared to the transition timescale $1/\beta$, as can be seen from the saddle point approximation of $P(t)$ around the percolation time t_p :

$$\begin{aligned} F(t) &\approx \exp(-0.34e^{\beta(t-t_p-\tau)}) - \exp(-0.34e^{\beta(t-t_p)}) \\ &= \beta\tau e^{\beta(t-t_p)} \exp(-0.34e^{\beta(t-t_p)}) \leq 0.37\beta\tau. \end{aligned} \quad (6.10)$$

Here, the last term follows by inserting the time at which $F(t)$ peaks, which is found to be $t \approx t_p - 1.08/\beta$. Alternatively, one can interpret F as the volume fraction of a shell around the bubbles with the width of the mean free path of the particles that just entered the bubbles. In the left panel of fig. 6.8 we show F as a function of $T - T_p$ for our benchmark point. As expected, we find that F takes its maximal value

*20 processes
contribute to the
DS thermalization*

*The
out-of-equilibrium
fraction $F(t)$...*

... is always small.

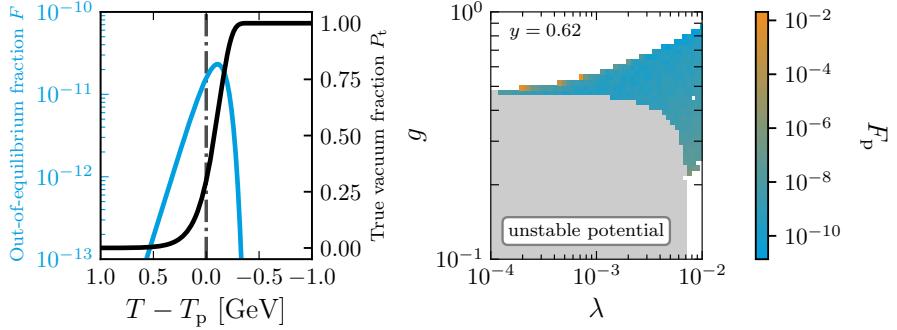


Figure 6.8: *Left:* The light blue line shows the energy fraction of the DS that is out of equilibrium for our benchmark point, as a function of the SM photon temperature. For reference, the fraction of the DS in the broken phase is given by the black line and the percolation temperature as the dashed vertical line. *Right:* Out-of-equilibrium fraction as a function of λ and g for $y = 0.62$ and $v_\phi = 430$ GeV.

close to the percolation temperature T_p . Nevertheless, even at the peak the value is tiny, implying that the fraction of the universe that is not in thermal equilibrium and therefore cannot be described by a temperature is completely negligible. In the right panel of fig. 6.8 we show the value $F_p \equiv F(T_p)$ at the percolation temperature for more points from our parameter scan, varying both g and λ . Similarly we find that for all our cases the thermalization is fast compared to the transition timescale. Even strongly supercooled points, where F_p becomes close to unity, are expected to return to equilibrium before freeze-out, since $\beta/H \gg 1$.

So far, we have throughout assumed that chemical potentials in the DS can be neglected after thermalization has taken place. This is certainly a good approximation as long as at least the lightest state in the DS has a mass that is small compared to the DS temperature, which is typically the case shortly after the PT. However, as the universe continues to cool down, this assumption becomes increasingly critical. If we assume that the DS cannot transfer its entropy to the SM thermal bath, the subsequent evolution depends crucially on whether number-changing processes of the lightest state, such as $3\phi \rightarrow 2\phi$, are efficient enough to maintain chemical equilibrium in the DS. If this is the case, the DS temperature will decrease much more slowly than the SM temperature, and the universe will enter a period of ‘cannibal’ domination [310, 311]. If, on the other hand, number-changing processes are inefficient, the DS will develop large chemical potentials and the universe will eventually enter a period of matter domination. In both cases, the energy and entropy stored in the DS must later be transferred to the SM heat bath in order to recover radiation domination before neutrinos decouple at $T \approx 2$ MeV, marking the onset of BBN [1]. Neither of these scenarios is very desirable, as the GW signals from the PT will be strongly diluted in the process [4].

*A note on
cannibalistic DSs*

We are therefore more interested in the case where the dark and SM sector quickly equilibrate after the PT, such that their temperatures

become equal, chemical potentials become negligible, and the universe evolves approximately as in radiation domination. In the following we discuss the processes that contribute to this process, and we derive the coupling strengths required for it to happen rapidly enough.

6.4.3 Thermalization of the dark and visible sector

The conceptually simplest way for the dark and visible sector to exchange entropy and energy is via Higgs mixing [298, 312–315].¹ Such a mixing arises from an additional term in the scalar potential:

$$V_{\text{mix}}(H, \Phi) = \lambda_{h\phi} |H|^2 |\Phi|^2. \quad (6.11)$$

As long as the vevs of both Higgs bosons vanish, the dominant process connecting the two sectors is $HH \rightarrow \Phi\Phi$. As soon as one of the two Higgs bosons acquires a vev, it can decay into the other one, i.e. $h \rightarrow \Phi\Phi$ (if the electroweak symmetry breaks first) or $\phi \rightarrow HH$ (if the dark symmetry breaks first). If kinematically allowed, these decay processes typically dominate over the $2 \rightarrow 2$ process for non-relativistic particles.

Thermalization through the Higgs portal

After both electroweak symmetry and the dark gauge symmetry have been spontaneously broken, the Higgs mixing generates a non-diagonal mass term, which can be rotated away by introducing the mixing angle

$$\theta = \frac{\lambda_{h\phi} v_\phi v_h}{m_h^2 - m_\phi^2}, \quad (6.12)$$

where we have assumed $\theta \ll 1$ both in order to satisfy experimental constraints on the properties of the SM-like Higgs boson and to ensure that thermal corrections from SM fields to the effective potential are negligible so that the DSPT can be treated separately from the EWPT. Both the masses and the vevs, and hence also the mixing angle, depend on the temperature. As a result of this mixing, the dark Higgs boson obtains couplings to SM fermions and gauge bosons proportional to θ . Of the greatest relevance for our discussion will be the decay of dark Higgs bosons into bottom quarks b , with a tree-level decay width given by

$$\Gamma_{\phi \rightarrow b\bar{b}} = \frac{3m_\phi m_b^2 \sin^2 \theta}{8\pi v_h^2} \sqrt{1 - \frac{4m_b^2}{m_\phi^2}}. \quad (6.13)$$

To calculate the entropy transfer between the dark and visible sector, we define the heat transfer rate

$$\dot{q}_{\text{DS}} \equiv \dot{\rho}_\phi + 3H(\rho_\phi + P_\phi), \quad (6.14)$$

¹ A second possibility would be to consider kinetic mixing between the dark photon and SM hypercharge. However, given that the dark photon is typically the heaviest state in the DS, it will be strongly Boltzmann-suppressed at low temperatures, and therefore cannot efficiently keep the two sectors in equilibrium.

which is related to the change of entropy density via

$$\dot{s}_{\text{DS}} = -3Hs_{\text{DS}} + \frac{\dot{q}_{\text{DS}}}{T_{\text{d}}}. \quad (6.15)$$

At the same time, the first moment of the Boltzmann equation gives

$$\int \frac{d^3p}{(2\pi)^3} E_\phi L[f_\phi] = \dot{q}_{\text{DS}} = \int \frac{d^3p}{(2\pi)^3} E_\phi C[f_\phi]. \quad (6.16)$$

The general expression for the first moment of the collision operator for decays (including relativistic corrections and quantum effects) was derived in refs. [316, 317]. It was shown there that the leading relativistic effects (namely a time dilation of the decay proportional to $1/\gamma$ and an increase of the injected energy by a factor γ) cancel, and it is therefore a good approximation to assume that the decaying particle is at rest. The integral of the collision operator thus gives

$$\dot{q}_{\text{DS}} \simeq m_\phi (\dot{n}_\phi + 3Hn_\phi), \quad (6.17)$$

and the evolution of the number density is given by the usual Boltzmann equation

$$\dot{n}_\phi + 3Hn_\phi = -\Gamma_{\phi \rightarrow b\bar{b}} n_\phi^{\text{eq}} \left(\frac{n_\phi}{n_\phi^{\text{eq}}} - 1 \right). \quad (6.18)$$

Here n_ϕ^{eq} denotes the equilibrium number density for $T_{\text{d}} = T$, whereas the actual number density can be calculated from the DS temperature and the assumption of negligible chemical potential. Putting everything together, we obtain

$$\dot{s}_{\text{DS}} = -3Hs_{\text{DS}} - \frac{m_\phi}{T_{\text{d}}} \Gamma_{\phi \rightarrow b\bar{b}} n_\phi^{\text{eq}} \left(\frac{n_\phi}{n_\phi^{\text{eq}}} - 1 \right). \quad (6.19)$$

A completely analogous equation holds for the evolution of the SM entropy density s_{SM} . Since the Hubble rate H depends on the combined energy density of both sectors, both equations need to be solved simultaneously, together with the equation $\dot{a} = Ha$, which we will need below to calculate the evolution of the GW spectrum.

*Entropy injection
to b quarks*

*We also include
other DS cooling
processes*

In practice, we also include decays into lighter quarks and leptons, which become relevant if decays into bottom quarks are kinematically forbidden. We further include the processes $h \rightarrow \phi\phi$ and $\phi \rightarrow hh$ if they are kinematically allowed. We do not, however, include $2 \rightarrow 2$ processes of the form $q\bar{q} \rightarrow g\phi$ or $q\phi \rightarrow qg$, which may give a non-negligible contribution for light dark Higgs bosons [318, 319]. Additional details, and the relevant equations, can be found in appendix 6.B. We note that analogous equations to the ones above can be derived for the case that only one Higgs boson has a vev and the case that both symmetries are unbroken.

As discussed above, it is not necessarily the case that the DS is in kinetic equilibrium with the SM at high temperatures. In the following, we will

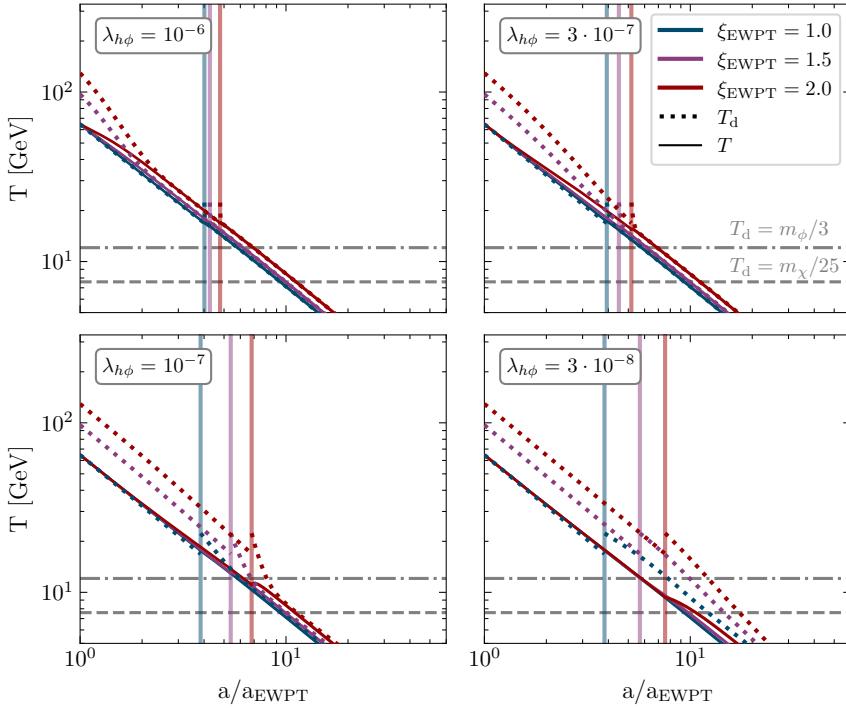


Figure 6.9: Evolution of the DS temperature (dotted) and the SM temperature (solid) for different values of the initial temperature ratio ξ_{EWPT} as a function of scale factor. The vertical lines indicate the scale factor at percolation, which depends on the temperature ratio. The horizontal lines indicate when the lightest state in the DS becomes non-relativistic ($T_d = m_\phi/3$) and approximately when the DM particles freeze out ($T_d = m_\chi/25$). The different panels correspond to different values of the portal coupling $\lambda_{h\phi}$.

therefore take the temperature ratio of the two sectors at the EWPT, ξ_{EWPT} , as a free parameter and calculate the evolution of ξ during the subsequent cosmological stages (see appendix 6.B for details). In fig. 6.9 we show the visible and DS temperatures as a function of the scale factor for different initial values of ξ_{EWPT} , defined at $T = 150$ GeV. In the four panels the portal coupling was set to the representative values $\lambda_{h\phi} = 10^{-6}$ (top left), $3 \cdot 10^{-7}$ (top right), 10^{-7} (bottom left) and $3 \cdot 10^{-8}$ (bottom right). In each panel, we indicate the moment of percolation by a vertical dot-dashed line; the approximate temperature of the dark Higgs becoming non-relativistic ($m_\phi/T_d = 3$) and the DM fermion freezing out ($m_\chi/T_d = 25$) are indicated by horizontal dotted and dashed lines, respectively. We make the following observations:

The portal coupling $\lambda_{h\phi}$ decides on the time of thermalization

- For $\lambda_{h\phi} = 10^{-6}$, the two sectors thermalize efficiently already before percolation. The initial value of ξ_{EWPT} is therefore inconsequential for the subsequent evolution, and we obtain the same results for all cases.
- For $\lambda_{h\phi} = 3 \times 10^{-7}$, the two sectors do exchange energy and entropy already in the unbroken phase, but do not fully ther-

malize before percolation, such that ξ_p depends on ξ_{EWPT} . After dark symmetry breaking, the two sectors thermalize rapidly, such that the subsequent evolution, and in particular the relic density calculation, do not depend on ξ_{EWPT} .

- For $\lambda_{h\phi} = 10^{-7}$, the energy exchange before dark symmetry breaking is completely negligible. Even after dark symmetry breaking, it will take a while for the temperatures of the two sectors to approach each other. Nevertheless, the two sectors reach equilibrium before the dark Higgs bosons become non-relativistic.
- For $\lambda_{h\phi} = 3 \times 10^{-8}$, the two sectors do not quickly thermalize after the PT, and the universe enters an early period of cannibal domination.²

All considered $\lambda_{h\phi}$ are well below direct detection constraints

In general the value of $\lambda_{h\phi}$ needed to ensure thermalization depends somewhat on the dark Higgs vev, since for $m_\phi \ll m_h$ the mixing angle scales as $\theta \propto \lambda_{h\phi} v_\phi$. Moreover, for small dark Higgs boson masses, decays into bottom quarks are kinematically forbidden and thermalization is less efficient. For the parameter points that reproduce the observed relic density we find that the assumption $\xi = 1$ made in section 6.3 is well justified for $\lambda_{h\phi}$ greater than 10^{-6} – 10^{-5} . This value should be compared to the currently strongest bounds from direct detection experiments, which are only sensitive to $\lambda_{h\phi} \gtrsim 10^{-3}$ [159].

6.5 GRAVITATIONAL WAVES FROM HOT DARK SECTORS

Turn up the volume: $\xi_p > 1$ enhances the GWs

While the analysis of the DSPT is simplest for $\lambda_{h\phi} > 10^{-6}$, it is phenomenologically interesting to also consider smaller values of $\lambda_{h\phi}$, such that the temperature ratio of the two sectors before the PT may differ from unity. The reason is that $\xi_p > 1$ leads to an enhancement of the GW signal, as a result of the larger total energy density in the DS compared to the SM thermal bath [4, 25]. In this case, however, we also need to consider what happens to the energy density of the DS after the PT.

If the transfer of energy from the dark to the visible sector after the PT is slow, the energy density of the universe will eventually be dominated by non-relativistic DS particles. This effect is already visible in fig. 6.9, where for small values of $\lambda_{h\phi} = 3 \times 10^{-8}$ the temperature ratio ξ still differs from unity when the lightest DS particle becomes non-relativistic. When the non-relativistic dark Higgs bosons eventually decay into SM particles, their entropy is transferred to the thermal bath; this modifies

² For the purpose of this plot, we assume that number-changing processes in the DS remain efficient throughout, such that the chemical potential of the dark Higgs boson vanishes and (in the absence of decays) the DS temperature grows relative to the SM temperature.

the expansion history of the universe and leads to a dilution of the GW signal. It is a-priori unclear whether this dilution effect dominates over the enhancement with increasing DS temperature, or whether a net increase in the strength of GW signals remains. Moreover, the dilution effect also shifts the GW frequencies and might thereby spoil the correlation between peak frequency and relic density found in section 6.3. In the following we will investigate these effects in detail.

If the portal coupling is extremely small, in principle even the relic density calculation could be modified. If the dark Higgs bosons become non-relativistic after freeze-out, in particular, they may come to dominate the energy density of the universe at later times, and dilute not only the GW energy density but also the DM energy density through their decays (see e.g. ref. [320]). If, on the other hand, the dark Higgs bosons are non-relativistic already during freeze-out, inefficient thermalization between the two sectors may additionally imply a non-trivial evolution of the DS temperature during freeze-out. While these are interesting scenarios in their own right, they are beyond the scope of this thesis. Instead, we will here focus exclusively on the case where the dark Higgs bosons decay sufficiently quickly for the standard freeze-out calculation (with temperature ratio $\xi = 1$) to be valid.

In section 4.3.2 we introduced the dilution factor $D \equiv S_{\text{SM}}^0/S_{\text{tot}}^{\text{p}}$ to account for the dilution of the GWB [4, 70]. In fig. 6.10 we show its dependence on the portal coupling $\lambda_{h\phi}$. Here, we choose the same benchmark point as studied in section 6.4, and show the result for different values of the initial DS temperature ratio ξ_{EWPT} , defined at $T = 150 \text{ GeV}$. As expected, for sufficiently large $\lambda_{h\phi}$ there is no significant dilution³, as entropy is conserved and only relativistic dofs contribute to the energy content of the universe. However, with decreasing $\lambda_{h\phi}$ this is no longer the case and the dilution factor grows, becoming sizable for $\lambda_{h\phi} \ll 10^{-7}$.

In fig. 6.11 we show the resulting GW spectra for $\xi \neq 1$ and $\lambda_{h\phi} < 10^{-6}$. As anticipated, a net enhancement of the GW signal is found for $\xi_{\text{EWPT}} > 1$, provided that $\lambda_{h\phi}$ is sufficiently small for the sectors to not equilibrate before the PT (cf. fig. 6.9). The enhancement saturates for $\xi_{\text{EWPT}} \gtrsim 2$, see also the discussion in ref. [4], implying a DS energy density that initially dominates over that of the SM sector. For too small portal couplings $\lambda_{h\phi} \lesssim 3 \times 10^{-8}$, on the other hand, the effect of dilution becomes relevant and the GW signal starts to become suppressed. Crucially, changing ξ and $\lambda_{h\phi}$ does not significantly affect the peak frequency, such that the GW signal remains within the LISA sensitivity range.

ξ_{p} and $\lambda_{h\phi}$ have competing effects

Other dilution effects

Dilution becomes relevant for $\lambda_{h\phi} \ll 10^{-7}$

A net enhancement remains for $\xi_{\text{p}} > 1$

³ In the limit of large portal couplings $\lambda_{h\phi}$, the dilution factor D approaches a value slightly larger than 1. This is an expected feature, indicating a negligible dilution effect that is entirely due to the additional degrees of freedom in the combined thermal bath of SM and DS particles and not a consequence of additional entropy injection (with respect to ΛCDM) into the SM bath after the PT.

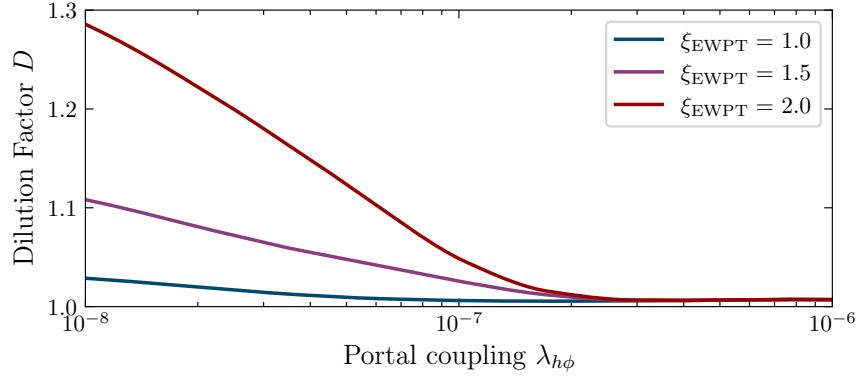


Figure 6.10: Dilution factor D , cf. eq. (4.35), as a function of the portal coupling $\lambda_{h\phi}$, and for different values of the initial temperature ratio ξ_{EWPT} as indicated.

We can now test the robustness of our results from section 6.3, where we assumed $\xi = 1$ and $\lambda_{h\phi} > 10^{-6}$, by allowing larger values of ξ and smaller values of $\lambda_{h\phi}$. In fig. 6.12 we show the same result as in fig. 6.6, but now for $\xi_{\text{EWPT}} = 2$ and $\lambda_{h\phi} = 10^{-7}$. For comparison we show the 1D distributions from fig. 6.6 as orange lines. In this plot we have removed points for which the DS temperature still differs significantly from the SM temperature when the dark Higgs boson becomes non-relativistic, i.e. for which $\xi_{\text{nr}} > 1.1$ at $T_{d,\text{nr}} = m_\phi/3$. The reason is that for such cases our final predictions depend on the details of chemical decoupling within the DS, which we do not study further in this thesis. While this requirement removes almost half of the parameter points, it does not introduce any significant bias, i.e. the distributions of f_{peak} and $\Omega_{\text{GW}}^{\text{peak}} h^2$ look very similar with and without this additional requirement.

The correlation is present also when $\xi_{\text{EWPT}} = 2$ and $\lambda_{h\phi} = 10^{-7}$

69% of points with strong supercooling are observable

The parameter combination $\xi_{\text{EWPT}} = 2$ and $\lambda_{h\phi} = 10^{-7}$ leads to a nearly maximal enhancement of the GW signal. As expected, we find that the peak position of the GW signal is not affected, such that the frequency range implied by the observed DM relic abundance remains within the LISA sensitivity window. The amplitude of the GW signal, on the other hand, is slightly enhanced, as can be seen from the comparison in the corresponding one-dimensional histogram. We can make this statement more precise by once again interpreting the density of points in the scatter plots as a probability distribution for the observables. Compared to fig. 6.6, we find that the probability to obtain a signal observable with LISA increases from 3% to 8%. Limiting ourselves to parameter regions with strong supercooling, the fraction of observable events increases from 35% to 69%. We summarize our findings in table 6.2.

We emphasize that this large increase is a result of fixing ξ and $\lambda_{h\phi}$ to particular values. If we instead vary ξ and $\lambda_{h\phi}$ as part of the scan, most parameter combinations will either give very similar results to the case $\xi = 1$ considered in section 6.3 or lead to an extended period where the DS energy density dominates.

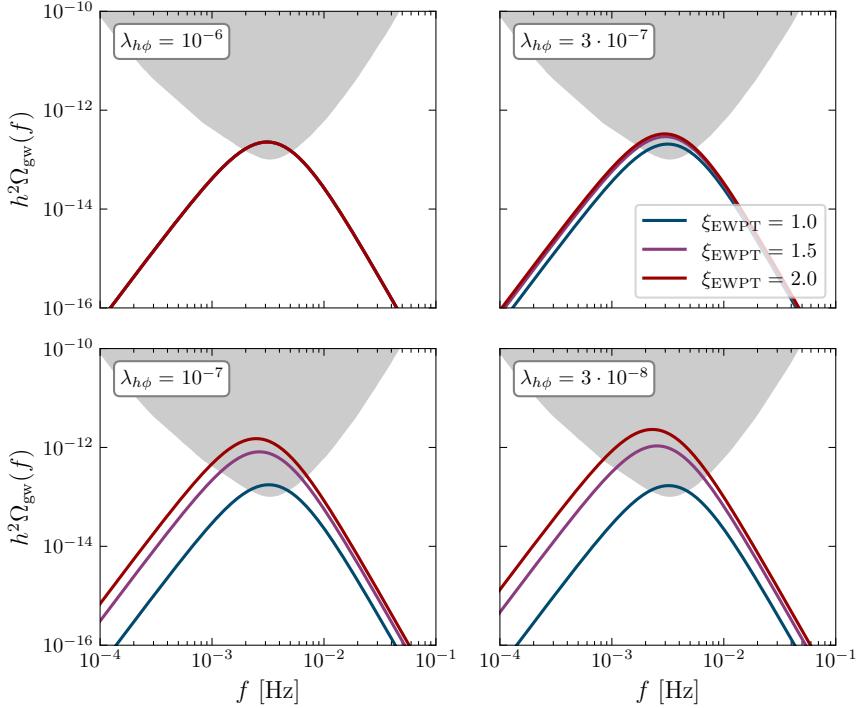


Figure 6.11: GW signal of the scenarios considered in fig. 6.9, i.e. for various values of the portal coupling $\lambda_{h\phi}$ (different panels) and the initial temperature ratio ξ_{EWPT} (different colors).

6.6 CONCLUSIONS

In this chapter we explored correlations between the DM relic density and GW signals arising from a FOPT that breaks a $U(1)'$ gauge symmetry and gives rise to the mass of a fermionic DM particle. We demonstrated that, while the amplitude of the GW signal depends on the details of the effective potential and can vary over many orders of magnitude, the peak frequency is tightly constrained once we impose the observed value for the DM relic abundance. Intriguingly, the peak frequency is found to lie exactly in the milli-Hertz range, which will be explored by the LISA mission.

A FOPT-triggered freeze-out hints towards LISA frequencies

The DS considered in this chapter is characterized by four parameters: the dark gauge coupling g , the quartic coupling of the dark Higgs field λ , the dark Yukawa coupling y and the dark Higgs vev v_ϕ . As a first step, we calculated the effective potential and the percolation temperature of the PT and identified the regions of parameter space that give a strong (large α) and not too fast ($\beta/H \sim 10^2\text{--}10^4$) PT, corresponding to potentially observable GW signals. We showed in particular that large GW signals require sizable couplings g and λ and occur also for large values of y . The relic density of the DS is determined by annihilations of DM fermions into pairs of dark Higgs bosons. The requirement to match the observed DM relic density then requires that the DM fermion

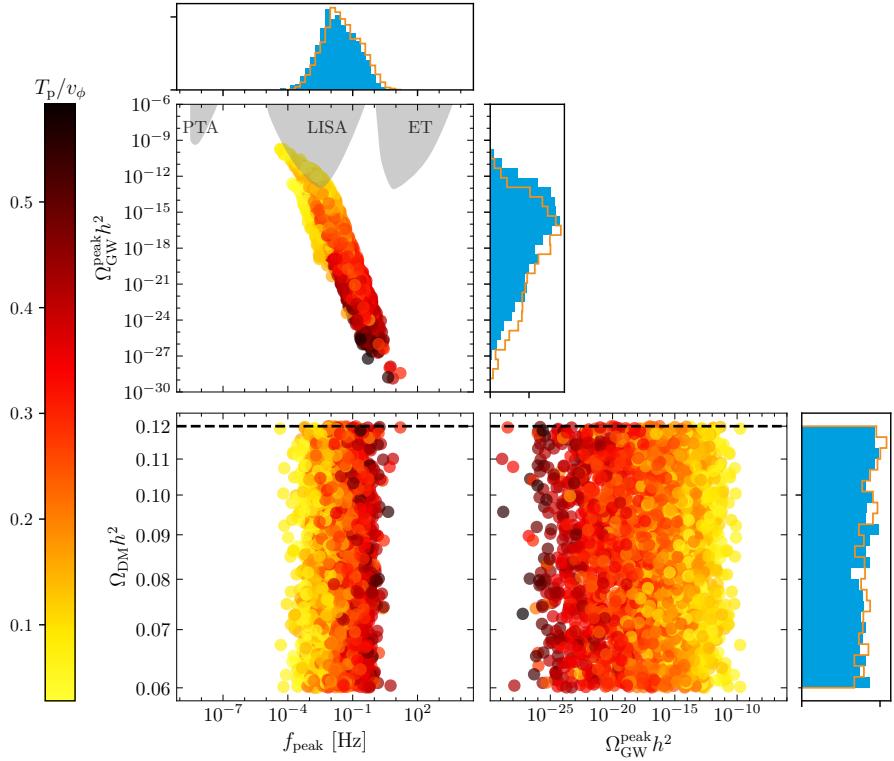


Figure 6.12: Same as fig. 6.6, but without the assumption of thermal equilibrium between the two sectors. Specifically, we consider $\xi_{\text{EWPT}} = 2$ and $\lambda_{h\phi} = 10^{-7}$. Compared to the situation in fig. 6.6 (indicated by the red lines in the 1D histograms), the GW amplitude is shifted to slightly larger values, while the peak position remains almost unaffected.

cannot be much lighter than the dark Higgs boson, with mass $m_\phi \sim v_\phi$, which in turn implies a sizable Yukawa coupling and a DM mass that is comparable to the dark Higgs vev v_ϕ . The dark Higgs vev, on the other hand, determines the percolation temperature and hence the peak frequency of the resulting GW signal. This connection leads to a tight correlation between relic density and GW peak frequency. Through comprehensive scans of the parameter space, we confirmed that this correlation is indeed highly generic in our model.

The correlation is due to requiring a strong FOPT & imposing $\Omega_{\text{DM}} h^2 = 0.12$

A FOPT producing relic DM is observable with 35% chance

A rigorous statistical interpretation of our results is beyond the scope of this chapter, but some estimates of the significance of the samples can be performed. A rough measure of the fine-tuning required to have a visible signal at LISA can be obtained by assuming that the sampling distributions of the parameters act as prior probabilities, and that the sampling density of points hence indicates the posterior distributions of derived quantities. Indeed, the majority of points drawn from these distributions do not feature any FOPT at all. Out of the points that feature a strongly-supercooled PT ($T_p/T_c < 0.5$), the probability of producing a visible signal at LISA in our model is around 10%. With the additional requirement that the observed DM relic abundance be

Selection requirement	Observable by LISA	
	$\xi_{\text{EWPT}} = 1,$ $\lambda_{h\phi} = 10^{-6}$	$\xi_{\text{EWPT}} = 2,$ $\lambda_{h\phi} = 10^{-7}$
Full sample	0.1%	0.5%
FOPT	0.8%	3%
FOPT + relic density	3%	8%
Strong supercooling	10%	21%
Strong supercooling + relic density	35%	69%

Table 6.2: Fraction of parameter points that predict an observable GW signal for LISA after imposing various selection requirements on the sample of points drawn from the parameter ranges discussed in section 6.2.3.

reproduced, this probability increases to 35%, as a result of the strong correlation between the predicted relic density and the peak frequency of the GW signal (see fig. 6.6).

We then studied two connected questions: How does the DS transfer its energy density to the SM? And is it justified to assume the same temperature for both sectors? Indeed, the PT leads to an increase of the DS temperature, as vacuum energy is converted into rest mass and kinetic energy. Having confirmed that the DS itself thermalizes immediately after the PT, we discussed in detail how the two sectors thermalize with each other for the specific case that the two Higgs bosons in the theory interact via the portal coupling $\lambda_{h\phi}$. After both electroweak symmetry breaking and DS symmetry breaking, this interaction leads to mixing between the Higgs bosons, such that they can decay into each other as well as into fermions of both sectors.

We derived and solved the Boltzmann equations for the entropy transfer between the two sectors and showed that for $\lambda_{h\phi} > 10^{-6}$ the assumption of equal temperature for both sectors is well justified. This portal coupling is small enough to be consistent with all laboratory constraints, in particular with direct detection experiments. For even smaller portal couplings, on the other hand, the temperatures of the two sectors may differ significantly, motivating us to consider the initial temperature ratio ξ_{EWPT} as a free parameter. While small values of $\lambda_{h\phi}$ combined with large initial values of ξ_{EWPT} may lead to an increase in the amplitude of the GW signal, the dark Higgs bosons will decay only slowly after the PT and may end up dominating the energy density of the universe. The resulting entropy injection would then lead to a substantial dilution of GW signals. We find that for $\lambda_{h\phi} \approx 10^{-7}$ a net enhancement remains, demonstrating that it is possible to have $\xi > 1$ during the DSPT while at the same time avoiding the dilution of the DS signal due to sufficiently rapid thermalization afterwards. Repeating our parameter scans for $\xi = 2$ and $\lambda_{h\phi} = 10^{-7}$, we find that the impact on the GW spectrum is only modest; while the typical amplitude is slightly enhanced,

*DS thermalization
is efficient*

*For $\lambda_{h\phi} \approx 10^{-7}$,
 $\xi > 1$ leads to
stronger GWs*

the peak frequency remains unchanged. In combination, these effects increase the fraction of points with a strongly supercooled PT that would be observable in LISA to 69%. Hence, our conclusion regarding the correlation between the DM relic abundance and the GW peak frequency applies also to DSs that thermalize only slowly with the SM.

An interesting open question is how the relic density calculation would change for even smaller values of the portal coupling than what we consider. In this case, the energy density of the universe would be dominated by non-relativistic dark Higgs bosons, which may develop a chemical potential if number-changing processes are inefficient. The relic density calculation then requires solving a coupled set of Boltzmann equations with non-trivial evolution of the DS temperature. While the details of this calculation are beyond the scope of this thesis, the general expectation is that the DM relic abundance would be increased. This might open up the possibility to have a DSPT in the nano-Hertz frequency range and hence of interest in the context of recent results from PTAs [7–9]. Another promising avenue for further investigations opens up due to our findings regarding the bubble wall velocity presented in appendix 6.A. There exists a potentially relevant part in the parameter space of our model in which the Bödeker-Moore criterion hints towards non-relativistic bubble wall velocities. For such low wall speeds the effect of bubble filtering [287] may be non-negligible, and the calculation of the DM relic abundance is expected to be more involved than presented here.

Let us finally mention that even for tiny portal couplings there is a chance to actually detect the DM particles that we consider: unlike for annihilation into dark Higgs boson pairs, the mixed annihilation channel into one dark Higgs boson and one dark photon is not suppressed in the limit of small DM velocities. If kinematically allowed, it may thus lead to observable signals in indirect detection experiments [321] such as CTA [322]. Such an observation would raise the possibility to explore in practice the correlations studied in this chapter and pin down the detailed structure of the DS.

*Indirect detection
probes*

*Can chemical
potentials push the
signal to nHz?*

6.A DARK BUBBLE WALLS

In addition to the strength α of the PT, its speed β/H , the percolation temperature T_p , and the temperature ratio ξ_p , it is also necessary to know the speed of the bubble walls v_w in order to determine the processes that dominate the GW signal from a DSPT. While the former parameters can be obtained from the effective potential $V_{\text{eff}}(\phi, T_d)$, the bubble wall velocity depends on plasma effects of the expanding bubble walls and therefore requires additional considerations. For bubbles expanding into the vacuum (i.e. if bubbles expand into a plasma that is not influenced by a change of the scalar vev), there is no source of friction, such that bubble walls can accelerate up to the point of their collision. For walls that interact with the surrounding plasma, on the other hand, several model-dependent sources of friction have been discussed [302, 323–326]. If the friction increases with the bubble wall velocity, the acceleration of the bubble walls eventually stalls and a terminal velocity is reached. In this case, the bulk motion of the plasma dominates the GW spectrum [304].

In this paper we take the approach suggested in ref. [208], i.e. we show that v_w is either expected to be non-relativistic in our model or that the bubbles are relativistic, but do not run away due to the emission of soft gauge bosons in the bubble walls. We conclude that the plasma motion is responsible for the dominant part of the GW signal and the contribution from bubble wall collisions is negligible.

To decide whether a bubble wall can accelerate up to relativistic velocities, we use the Bödeker-Moore criterion [324], which relates the velocity-independent leading-order (LO) bubble wall friction \mathcal{P}_{LO} to the amount of liberated vacuum energy density ΔV_{eff} :

$$\text{BM criterion: } \begin{cases} \Delta V_{\text{eff}} > \mathcal{P}_{\text{LO}} & \text{Relativistic bubble walls.} \\ \Delta V_{\text{eff}} < \mathcal{P}_{\text{LO}} & \text{Non-relativistic bubble walls.} \end{cases} \quad (6.20)$$

We emphasize that this criterion is insufficient to decide whether walls can run away (i.e. accelerate indefinitely), because of the next-to-leading-order (NLO) friction \mathcal{P}_{NLO} , which scales with powers of $\gamma_w = 1/\sqrt{1-v_w^2}$ [302, 303, 327]. Bubbles can run away only if $\Delta V_{\text{eff}} > \mathcal{P}_{\text{LO}} + \mathcal{P}_{\text{NLO}}$ for all v_w . Otherwise they will reach a relativistic terminal velocity given by the equilibrium of forces, $\Delta V_{\text{eff}} = \mathcal{P}_{\text{LO}} + \mathcal{P}_{\text{NLO}}$.

The LO friction due to particles acquiring a mass when crossing the bubble wall is given by [192, 324]

$$\mathcal{P}_{\text{LO}} \simeq \sum_i g_i c_i \frac{\Delta m_i^2}{24} T_{d,p}^2, \quad (6.21)$$

where Δm_i^2 is the positive change of the mass square of particle species i during the PT, g_i is the corresponding number of degrees of freedom, and $c_i = 1$ ($1/2$) for bosons (fermions). This expression assumes that

The dark wall velocity

Gauge bosons prohibit runaway

The Bödeker-Moore criterion

LO friction from mass increase Δm_i^2

the particle masses outside the bubble are below the DS temperature. For ultra-relativistic bubble walls, the production of heavy particles with mass up to $\gamma_w T_{d,p}$ can also add to the LO friction [325]. For the DS considered in this chapter, all particles are light before the PT and subsequently obtain a mass comparable to the scale of the PT, such that (6.21) is the only relevant contribution to the LO friction.

In the case of a broken $U(1)'$ with a fermionic species, the LO friction reads [192]

$$\mathcal{P}_{\text{LO}} \simeq (3m_{A'}^2 + m_\phi^2 + 2m_\chi^2) \frac{T_{d,p}^2}{24} \approx \frac{m_{A'}^2 T_{d,p}^2}{8}, \quad (6.22)$$

where in the last step we have used that the dominant contribution comes from the heaviest state in the DS, which in the parameter regions of interest is the dark photon. The amount of released vacuum energy ΔV_{eff} can be estimated from the zero-temperature potential, which gives

$$\Delta V_{\text{eff}} \approx \frac{\lambda}{4} v_\phi^4 = \frac{m_\phi^2}{8} v_\phi^2. \quad (6.23)$$

Hence we find that the Bödeker-Moore criterion for relativistic bubble walls is satisfied if $m_\phi/m_{A'} > T_{d,p}/v_\phi$, which is the case for the parameter regions that give strongly supercooled PTs corresponding to observable GW signals ($T_{d,p} \lesssim 0.1 v_\phi$). In these parameter regions, the bubble walls are therefore expected to be relativistic, $v_w \rightarrow 1$. This finding also implies that we can neglect the effect of bubble filtering, which is only relevant in the (deeply) non-relativistic regime $v_w \ll 1$ [287, 309]. In the regions in which weaker GW signals are expected, the Bödeker-Moore criterion instead hints towards slower bubble walls, see fig. 6.13.

The NLO friction created by the emission of soft dark photons into the broken phase quickly starts to grow with γ_w . The bubble walls will therefore reach a terminal, asymptotic bubble wall velocity which is close to the speed of light. The precise value of $\gamma_w^{\text{terminal}}$ is unnecessary for our purposes, as the existence of a terminal yet relativistic bubble wall velocity is sufficient to assume a dominant GW emission through bulk fluid motion. A more refined calculation of the respective energy budgets for the processes emitting gravitational radiation was performed in refs. [185, 328, 329]. In ref. [328] it was shown that for sufficiently high terminal bubble wall velocities the fluid profiles are strongly peaked, such that the emitted GW spectral shapes are in fact indistinguishable from bubble collisions. We conclude that it is hence a reasonable approximation to work with a GW spectrum that is solely sourced through sound waves.

*Strong supercooling
implies fast walls*

*Sound waves are
dominant GW
source*

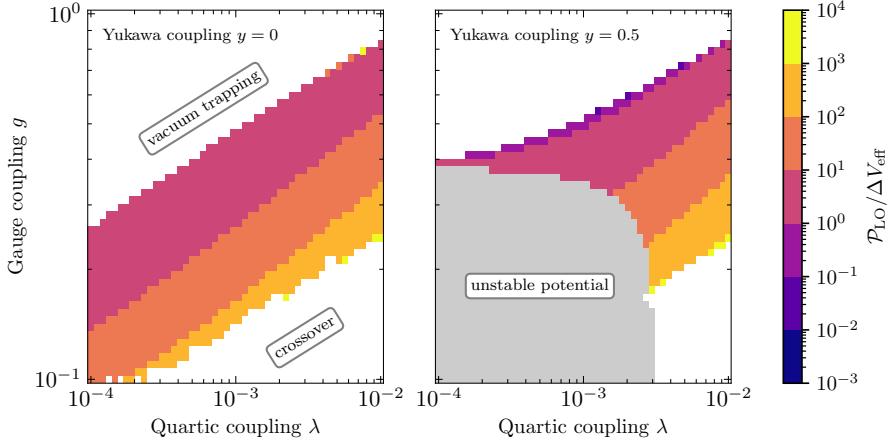


Figure 6.13: The leading-order friction \mathcal{P}_{LO} over the difference ΔV_{eff} in potential energy between the true and false vacuum phases as a function of the quartic coupling λ and the gauge coupling g for values of the Yukawa coupling $y = 0$ and $y = 0.5$. Values greater than one indicate that relativistic bubble wall velocities cannot be reached, cf. eq. (6.20). For values smaller than one, a relativistic terminal velocity is expected.

6.B DETAIL ON THERMALIZATION AND FREEZE-OUT

Here we discuss the processes we consider for the entropy transfer between the DS and the SM.

6.B.1 The thermal mixing angle

The Higgs mixing angle defined in eq. (6.12) depends on temperature through the masses and vevs of the two Higgs bosons. The temperature dependence of the dark Higgs boson can be directly obtained from the effective potential, whereas we follow ref. [316] to implement the temperature dependence of the SM Higgs boson. For large values of the dark Higgs vev, we sometimes encounter the situation that the SM Higgs and dark Higgs mass become similar or even cross, in which case the mixing angle apparently diverges. To regulate this unphysical divergence, we have to include the finite width Γ_h of the Higgs resonance. As shown in ref. [319], including the width leads to an effective mixing angle given by

$$\theta_{\text{eff}}^2(T) = \frac{(\lambda_{h\phi} v_h(T) v_\phi(T))^2}{(m_\phi^2(T) - m_h^2(T))^2 + (m_\phi(T) \Gamma_h)^2}. \quad (6.24)$$

Mixing depends on temperature

6.B.2 The Boltzmann equation for entropy transfer

On the order of DSPT and EWPT

In our analysis we specify the initial conditions, i.e. the temperature ratio ξ_{EWPT} of the DS and SM bath at the EWPT, for which we take $T_{\text{EWPT}} = 150 \text{ GeV}$. Which processes contribute to the thermalization of the two sectors depends on whether or not the $U(1)'$ gauge symmetry is already broken at this point. We consider two different timelines:

- The electroweak symmetry breaks before the DSPT. This is the case for the majority of our parameter space. Here, the thermalization between the two sectors is initially determined by the processes $hh \leftrightarrow \Phi\Phi$ and the decay of the SM Higgs $h \rightarrow \Phi\Phi$. Additional processes can only contribute after the DSPT.
- The DSPT occurs before the EWPT. This can happen for parameter points with a large vev of the dark Higgs and not too strong supercooling. Once both symmetries are broken, thermalization proceeds via $hh \leftrightarrow \phi\phi$, the decays of the SM Higgs into dark Higgs $h \rightarrow \phi\phi$ if it is kinematically allowed, and the decays of the dark Higgs into SM particles $\phi \rightarrow \text{SM SM}$ through Higgs mixing.

In the following we give the relevant expressions contributing to the entropy transfer.

Thermalization through $\Phi\Phi \leftrightarrow hh$

$2 \rightarrow 2$ PROCESSES In most regions of parameter space, the DS phase is unbroken immediately after the EWPT, such that there are no dark Higgs boson decays that can transfer entropy between the two sectors. Here the $2 \rightarrow 2$ process induced by the portal coupling $\Phi\Phi \leftrightarrow hh$ become relevant. Since the particles' thermal masses are smaller than the temperature, we cannot take the usual non-relativistic approach. Instead we will follow the relativistic treatment of the calculation of the entropy transfer developed in refs. [330, 331], which we briefly sketch here. The heat transfer for $2 \rightarrow 2$ processes can be expressed as

$$\begin{aligned} \dot{q}|_{2 \rightarrow 2}^{\text{DS}} &= \int \frac{d^3 p_1}{(2\pi)^3} \frac{d^3 p_2}{(2\pi)^3 2E_2} \frac{d^3 k_1}{(2\pi)^3 2E_{k_1}} \frac{d^3 k_2}{(2\pi)^3 2E_{k_2}} |\mathcal{M}|^2 (2\pi)^4 \delta(\Sigma p) \\ &\quad \times f(p_1) f(p_2) (1 + f(k_1)) (1 + f(k_2)) \\ &= \int \frac{d^3 p_1}{(2\pi)^3 2E_1} \frac{d^3 p_2}{(2\pi)^3 2E_2} 8E_1 F(p_1, p_2) \sigma(p_1, p_2) f(p_1) f(p_2). \end{aligned} \quad (6.25)$$

Here the final state statistical factors are absorbed into the cross section and $F(p_1, p_2) = \sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}$. It is easiest to calculate this cross section in the center of mass frame. However, since the Bose-Einstein and Fermi-Dirac distributions are not Lorentz-invariant we have to apply a Lorentz boost Λ from the cosmic rest frame where $u = (1, 0, 0, 0)^T$ into the center of mass frame, which can be parameterized by the rapidity

η and two angles θ and φ ; for details see ref. [330]. The phase space distribution becomes

$$f(k) = \frac{1}{e^{u \cdot k/T_d} \mp 1} \xrightarrow{\Lambda} f_\Lambda(k) = \frac{1}{e^{(k_0 \cosh \eta + k_1 \sinh \eta)/T_d} \mp 1}. \quad (6.26)$$

With this, we can rewrite the center of mass cross section as

$$\begin{aligned} \sigma_{\text{CM}}(p_1, p_2) &= \frac{1}{(2\pi)^2 16F(p_1, p_2)} \int d\varphi d\cos\theta |\mathcal{M}|^2 \\ &\times \frac{\sqrt{E^2 - m_f^2}}{E} ((1 + f_\Lambda(k_1))(1 + f_\Lambda(k_2))). \end{aligned} \quad (6.27)$$

The matrix element for the processes $\Phi\Phi \rightarrow hh$ is at tree level simply given by $\mathcal{M} = i\lambda_{h\phi}$. For this angle-independent transition amplitude, we can integrate over the solid angle and obtain

$$\begin{aligned} \sigma_{\text{CM}} &= \frac{|\mathcal{M}|^2}{64\pi E^2} \frac{T_d}{\sqrt{E^2 - m_\Phi^2} \sinh \eta} \frac{1}{1 - e^{-\frac{2E}{T_d} \cosh \eta}} \\ &\times \ln \underbrace{\left[\frac{\sinh \left(\frac{E \cosh \eta + \sqrt{E^2 - m_h^2} \sinh \eta}{2T_d} \right)}{\sinh \left(\frac{E \cosh \eta - \sqrt{E^2 - m_h^2} \sinh \eta}{2T_d} \right)} \right]}_{\equiv \lambda(E, \eta, T_d, m_h)}. \end{aligned} \quad (6.28)$$

In the case of initial and final states with respectively equal masses, eq. (6.25) reduces to

$$\begin{aligned} I_{\Phi\Phi \rightarrow hh} &\equiv \frac{2T_d}{\pi^4} \int_{m_\Phi(T_d)}^\infty dE \sqrt{E^2 - m_\Phi^2} E^4 \\ &\times \int_0^\infty d\eta \frac{\sigma_{\text{CM}} \sinh \eta \cosh \eta}{e^{2E \cosh \eta / T_d} - 1} \ln \lambda(E, \eta, T_d, m_\Phi). \end{aligned} \quad (6.29)$$

This expression can now be efficiently evaluated numerically. An analogous expression is obtained for the process $hh \rightarrow \phi\phi$.

We note that in principle there are additional $2 \rightarrow 2$ processes that may contribute to thermalization. The process $\phi\phi \rightarrow t\bar{t}$ via an off-shell SM Higgs boson is strongly Boltzmann-suppressed below the EWPT [316] and does not give a relevant contribution in the temperature range that we consider. However, processes such as $\phi + q \rightarrow g + q$ with a quark in the t -channel can give a relevant contribution if the decay $\phi \rightarrow b\bar{b}$ is kinematically forbidden [318, 319]. Since this is the case only in a very small fraction of the parameter space that we consider, we neglect these processes, thus giving a conservative estimate of the thermalization rate.

We neglect
 $\phi + q \rightarrow g + q$

STANDARD MODEL HIGGS BOSON DECAYS After both symmetries are broken and for temperatures comparable to the SM Higgs boson mass,

a second process of interest besides dark Higgs decays is the resonantly enhanced pair-annihilation of dark Higgs bosons into predominantly bottom quarks: $\phi\phi \rightarrow h \rightarrow b\bar{b}$. In thermal equilibrium, the rate of this process can be related to the inverse process, which is the decay $h \rightarrow \phi\phi$ with partial width given by

$$\Gamma_{h \rightarrow \phi\phi} = \frac{(m_h^2 + 2m_\phi^2)^2 \sin^2 2\theta_{\text{eff}}}{128\pi m_h} \left(1 - \frac{4m_\phi^2}{m_h^2}\right)^{1/2} \times \left(\frac{1}{v_\phi} \cos \theta_{\text{eff}} + \frac{1}{v_h} \sin \theta_{\text{eff}}\right)^2. \quad (6.30)$$

This gives the additional term in the heat transfer rate

$$\dot{q}|_{h \rightarrow \phi\phi}^{\text{DS}} = -m_h \Gamma_{h \rightarrow \phi\phi} n_h^{\text{eq}} \left[1 - \left(\frac{n_\phi}{n_\phi^{\text{eq}}}\right)^2\right]. \quad (6.31)$$

Before the DSPT, we also have the decay $h \rightarrow \Phi\Phi$ with the decay rate

$$\Gamma_{h \rightarrow \Phi\Phi} = \frac{\lambda_{h\phi}^2 v_h^2}{128\pi m_h}, \quad (6.32)$$

which can be treated in analogy to the case above.

FULL BOLTZMANN EQUATION The full Boltzmann equation that we solve before the PT, in case it occurs after the EWPT then reads

$$\begin{aligned} \dot{s}_{\text{DS}} + 3Hs_{\text{DS}} &= -\frac{m_h}{T_d} \Gamma_{h \rightarrow \Phi\Phi} n_h^{\text{eq}} \left[\left(\frac{n_\phi}{n_\phi^{\text{eq}}}\right)^2 - 1\right] \\ &\quad - \frac{1}{T_d} I_{\Phi\Phi \rightarrow hh} + \frac{1}{T_d} I_{hh \rightarrow \Phi\Phi}. \end{aligned} \quad (6.33)$$

*Before the DSPT,
after the EWPT*

The equation for the case of the DSPT occurring before the EWPT follows analogously. After both symmetries are broken the full Boltzmann equation reads

$$\begin{aligned} \dot{s}_{\text{DS}} + 3Hs_{\text{DS}} &= -\frac{m_\phi}{T_d} \Gamma_{\phi \rightarrow \text{SMSM}} n_\phi^{\text{eq}} \left(\frac{n_\phi}{n_\phi^{\text{eq}}} - 1\right) \\ &\quad + \frac{m_h}{T_d} \Gamma_{h \rightarrow \phi\phi} n_h^{\text{eq}} \left[1 - \left(\frac{n_\phi}{n_\phi^{\text{eq}}}\right)^2\right] \\ &\quad - \frac{1}{T_d} I_{\phi\phi \rightarrow hh} + \frac{1}{T_d} I_{hh \rightarrow \phi\phi}. \end{aligned} \quad (6.34)$$

After both PTs

where we use the decay widths of the SM Higgs into other SM particle pairs from ref. [159]. The equations for the SM entropy follow analogously.

6.B.3 Annihilation cross sections

In the following we list the various DM annihilation cross sections in the non-relativistic limit, up to second order in the center-of-mass velocity v (of each of the DM particles):

$$\begin{aligned}
 (\sigma v)_{\chi\chi \rightarrow A'A'} = & \frac{m_{A'}^4(m_\chi^2 - m_{A'}^2)^{3/2}}{64\pi v_\phi^4 m_\chi(m_{A'}^2 - 2m_\chi^2)^2} \\
 & + \frac{v^2 \sqrt{m_\chi^2 - m_{A'}^2}}{384\pi v_\phi^4 m_\chi(m_\phi^2 - 4m_\chi^2)^2(m_{A'}^2 - 2m_\chi^2)^4} \\
 & \times \left[144m_{A'}^{12}m_\chi^2 \right. \\
 & + 2m_{A'}^{10}(7m_\phi^4 - 88m_\phi^2m_\chi^2 - 432m_\chi^4) \\
 & + 128m_\chi^{10}(m_\phi^4 + 8m_\chi^4) \\
 & - 64m_{A'}^2m_\chi^8(3m_\phi^4 + 16m_\phi^2m_\chi^2 + 32m_\chi^4) \\
 & + 4m_{A'}^4m_\chi^6(17m_\phi^4 + 600m_\phi^2m_\chi^2 + 592m_\chi^4) \\
 & + m_{A'}^8(-73m_\phi^4m_\chi^2 + 1128m_\phi^2m_\chi^4 + 1840m_\chi^6) \\
 & \left. + 4m_{A'}^6(25m_\phi^4m_\chi^4 - 648m_\phi^2m_\chi^6 - 496m_\chi^8) \right] \\
 & + \mathcal{O}(v^4)
 \end{aligned} \tag{6.35}$$

$$\begin{aligned}
 (\sigma v)_{\chi\chi \rightarrow \phi\phi} = & \frac{v^2 m_\chi \sqrt{m_\chi^2 - m_\phi^2} (3m_\phi^4 - 8m_\phi^2m_\chi^2 + 8m_\chi^4)}{192\pi v_\phi^4 (m_\phi^2 - 4m_\chi^2)^2 (m_\phi^2 - 2m_\chi^2)^4} \\
 & \times (9m_\phi^8 - 64m_\phi^6m_\chi^2 + 200m_\phi^4m_\chi^4 \\
 & - 352m_\phi^2m_\chi^6 + 288m_\chi^8) \\
 & + \mathcal{O}(v^4)
 \end{aligned} \tag{6.36}$$

$$\begin{aligned}
 (\sigma v)_{\chi\chi \rightarrow A'\phi} = & \frac{1}{2048\pi v^4 m_\chi^4} \\
 & \times (m_{A'}^4 - 2m_{A'}^2m_\phi^2 + m_\phi^4 - 8m_{A'}^2m_\chi^2 \\
 & - 8m_\phi^2m_\chi^2 + 16m_\chi^4)^{3/2} \\
 & + \frac{v^2 \sqrt{m_{A'}^4 + (m_\phi^2 - 4m_\chi^2)^2 - 2m_{A'}^2(m_\phi^2 + 4m_\chi^2)}}{12288\pi v_\phi^4 m_\chi^4 (m_{A'}^2 - 4m_\chi^2)^2 (m_{A'}^2 + m_\phi^2 - 4m_\chi^2)^4} \\
 & \times \left[-11m_{A'}^{16} - 2m_{A'}^{14}(11m_\phi^2 - 228m_\chi^2) \right. \\
 & - 16m_\chi^4(m_\phi^2 - 4m_\chi^2)^4 \\
 & \times (15m_\phi^4 - 80m_\phi^2m_\chi^2 + 16m_\chi^4) \\
 & - m_{A'}^{12}(-11m_\phi^4 - 1200m_\phi^2m_\chi^2 + 7264m_\chi^4) \\
 & + 4m_{A'}^{10}(11m_\phi^6 + 266m_\phi^4m_\chi^2 \\
 & \left. - 4328m_\phi^2m_\chi^4 + 15136m_\chi^6) \right]
 \end{aligned}$$

$$\begin{aligned}
& + 8m_{A'}^2 m_\chi^2 (m_\phi^2 - 4m_\chi^2)^2 \\
& \times (15m_\phi^8 - 244m_\phi^6 m_\chi^2 + 1280m_\phi^4 m_\chi^4 \\
& \quad - 2880m_\phi^2 m_\chi^6 + 6912m_\chi^8) \\
& - m_{A'}^4 (m_\phi^2 - 4m_\chi^2)^2 (11m_\phi^8 - 344m_\phi^6 m_\chi^2 \\
& \quad + 1808m_\phi^4 m_\chi^4 - 8704m_\phi^2 m_\chi^6 + 80384m_\chi^8) \\
& - m_{A'}^8 (-11m_\phi^8 - 32m_\phi^6 m_\chi^2 + 12464m_\phi^4 m_\chi^4 \\
& \quad - 114688m_\phi^2 m_\chi^6 + 291840m_\chi^8) \\
& - 2m_{A'}^6 (11m_\phi^{10} - 140m_\phi^8 m_\chi^2 + 1216m_\phi^6 m_\chi^4 \\
& \quad - 29056m_\phi^4 m_\chi^6 + 201984m_\phi^2 m_\chi^8 \\
& \quad - 412672m_\chi^{10}) \\
& + \mathcal{O}(v^4)
\end{aligned} \tag{6.37}$$

7

DO PTAS OBSERVE INSPIRALING PRIMORDIAL BLACK HOLES?

This chapter is based on the following publication:

- [3] P. F. Depta, K. Schmidt-Hoberg, P. Schwaller, and C. Tasillo,
Do pulsar timing arrays observe merging primordial black holes?, [[2306.17836](#)]

*Our hopes and expectations:
 Black holes and revelations*

— *Starlight by MUSE*

7.1 INTRODUCTION

In chapter 3 we reviewed the recent PTA results and concluded that the inspiral of SMBHBs is expected to contribute to the observed GWB at nHz frequencies [238]. To match the observed signal amplitude, however, the local SMBHB density would need to be an $\mathcal{O}(10)$ factor larger than previously estimated [216–218]. Hence, the validity of this explanation is under active debate [219–222] and it is interesting to consider alternative sources.

In chapter 5 we studied whether cosmological PTs could account for the novel PTA signal. In the present chapter we move on and study the possibility that the observed GWB could be due to inspiraling supermassive *primordial* black holes. The existence of PBHs in this mass range is motivated by yet unresolved puzzles concerning the observation of a large population of high-redshift quasars [29–33] and the related problem of missing seeds for non-linear structure formation at early times [34, 332].

To obtain the correct GWB signal amplitude, however, we will find that a sizable abundance of PBHs is required, which is subject to strong constraints from various observational probes [28]. It is therefore a non-trivial question whether inspiraling PBHs could constitute a viable explanation. Indeed we find, in accordance with ref. [123] which appeared shortly after our work [3], that it is *not possible* to consistently explain the NANOGrav data with homogeneously distributed PBHs. Specifically the parameter regions at very large PBH masses which naively allow fitting the PTA data (as done in [333]) no longer result in a stochastic

*Astrophysical
 SMBHBs can only
 partially explain
 NANOGrav*

*Motivation for
 supermassive
 PBHs*

*Our results
 contradict Atal et
 al.*

Outline of this chapter

Going beyond
 $\Omega_{\text{gw}}(f) \propto f^{2/3}$

(or even continuous) signal. For clustered PBHs, on the other hand, we find that the formation of binary systems is more efficient and the merger rate is generally enhanced [334], opening the possibility for an overall consistent description of the observed GWB.

This chapter is structured as follows: We first review and extend the formalism required to compute the GWB emitted from binary mergers for the case of clustered PBHs in section 7.2. In section 7.3 we summarize our calculation of the number of PBH binaries responsible for a given GWB, justifying our claim that ref. [333] came to an erroneous result. After discussing the relevant PBH constraints and how to match the predicted signal to PTA data in sections 7.4 and 7.5, respectively, we present our results in section 7.6.

7.2 GRAVITATIONAL WAVE SIGNAL

We provided a first approximation for computing the GWB spectrum from the inspiral of a population of SMBHBs in eq. (3.31). Now, we will go beyond this: In the scenario at hand, the GWB $h^2\Omega_{\text{gw}}(f)$ is generated from PBH binary mergers occurring at a rate $R(t_r)$ per comoving volume V_c and cosmic time t_r . The frequencies accessible to PTAs can be well below the maximal frequency emitted during the merger and might therefore have been radiated long before coalescence, which we need to take into account. In particular, a given binary merging at time $t_{r,\text{merg}}$ has emitted a frequency f_r in the cosmic rest frame, redshifted to a frequency $f = f_r/(1+z)$ today, at time $t_r = t_{r,\text{merg}} - \tau_{f_r}$, where

$$\tau_{f_r} = \frac{5 \times 2^{1/3}}{256 \pi^{8/3}} f_r^{-8/3} (Gm_{\text{PBH}})^{-5/3} \quad (7.1)$$

is the time until coalescence for f_r [14]. As $t_{r,\text{merg}}$ is the argument for the *merger* rate, the number of events per comoving volume and cosmic time emitting a frequency f_r in the cosmic rest frame at cosmic time t_r is given by $R(t_r + \tau_{f_r})$. Note that to the best of our knowledge this frequency-dependent rate has not been discussed in the literature concerning GWs generated by PBHs before. The GW energy density parameter is given by (cf. ref. [150])

$$\Omega_{\text{gw}}(f) = \frac{f}{\rho_{\text{crit}}} \int_0^{t_0} dt_r \left(R(t_r + \tau_{f_r}) \frac{dE_{\text{gw}}^r}{df_r} \right)_{f_r=(1+z)f} , \quad (7.2)$$

where t_0 is the current cosmic time, $f_r = (1+z)f$ is the frequency that needs to be emitted at the redshift $z = z(t_r)$ corresponding to t_r to

detect a frequency f today, and the GW spectrum can be estimated by power-laws for inspiral, merger, and ringdown [335]¹

$$\frac{dE_{\text{gw}}^r}{df_r} \sim \frac{(\pi G)^{2/3} m_{\text{PBH}}^{5/3}}{3 \times 2^{1/3}} \begin{cases} f_r^{-1/3} & f_r < f_1 \\ \frac{f_r^{2/3}}{f_1} & f_1 \leq f_r < f_2 \\ \frac{f_r^2 f_4^4}{f_1 f_2^{4/3} [4(f_r - f_2)^2 + f_4^2]^2} & f_2 \leq f_r < f_3 \\ 0 & f_3 \leq f_r . \end{cases} \quad (7.3)$$

Here we assumed PBHs with equal masses m_{PBH} , $f_i = (a_i \eta^2 + b_i \eta + c_i) / (2\pi G m_{\text{PBH}})$ with $\eta = 1/4$, and used the coefficients provided in Table I of ref. [335]

$$(a_1, a_2, a_3, a_4) = (2.97, 5.94, 8.48, 5.08) \times 10^{-1}, \quad (7.4a)$$

$$(b_1, b_2, b_3, b_4) = (4.48, 8.98, 12.8, 7.75) \times 10^{-2}, \quad (7.4b)$$

$$(c_1, c_2, c_3, c_4) = (9.56, 19.1, 27.3, 2.24) \times 10^{-2}. \quad (7.4c)$$

GWs from inspiral, merger and ringdown

Note that the GW energy density $\Omega_{\text{gw}}(f)$ is in fact a cosmological average of many PBH binaries. We discuss the effect of only having access to a local realization of the GWB within PTA measurements in sec. 7.3.

For the merger rate we adapt the calculation from ref. [336]. We assume a monochromatic initial PBH mass distribution, expecting that going to an extended distribution does not qualitatively change our results. A PBH binary forms in the early Universe when the gravitational attraction between two neighboring PBHs overcomes the Hubble flow. A third close-by PBH provides angular momentum such that the two PBHs do not simply collide [337–339].

Clearly the merger rate depends on the local PBH density at binary formation. It is therefore interesting to study the effect of clustering, which increases the global comoving number density n_{PBH} by the local density contrast δ_{dc} . The density contrast can be considered constant on the scales relevant for binary formation [336]. The comoving number density n_{PBH} can be expressed in terms of the PBH mass m_{PBH} , the fraction of PBH DM f_{PBH} and the DM density $\rho_{\text{DM},0}$ today

$$n_{\text{PBH}} = f_{\text{PBH}} \frac{\rho_{\text{DM},0}}{m_{\text{PBH}}}. \quad (7.5)$$

Clustering δ_{dc} increases the merger rate

The number density dn_3 of the three-body configurations relevant for binary formation, where the comoving distance from a given PBH in the binary to the other PBH in the binary is within x and $x + dx$ and the comoving distance to the third PBH is within y and $y + dy$, is [336]

$$dn_3(x, y) = \frac{n_{\text{PBH}}}{2} e^{-N_{\text{PBH}}(y)} (4\pi n_{\text{PBH}} \delta_{\text{dc}})^2 x^2 y^2 dx dy. \quad (7.6)$$

¹ In our calculations we include a cut-off at $f_r = 1/t_r$, since smaller frequencies cannot have completed even one complete orbit.

Here, the factor of $1/2$ removes over-counting as either one of the PBHs in the binaries can be chosen and $e^{-N(y)}$ makes sure that there is no fourth binary within the comoving distance y (assuming Poisson statistics), where

$$N_{\text{PBH}}(y) = \frac{4\pi}{3} n_{\text{PBH}} \delta_{\text{dc}} y^3 \quad (7.7)$$

is the expected number of binaries within a sphere of comoving distance y . The coalescence time τ of this binary is given by [336, 340]

$$\tau(x, y) = \tilde{\tau} \left(\frac{x}{\tilde{x}} \right)^{37} \left(\frac{y}{\tilde{x}} \right)^{-21}, \quad (7.8)$$

where

$$\tilde{\tau} = \frac{3}{170} \frac{(a_{\text{eq}} \tilde{x})^4}{(G m_{\text{PBH}})^3}, \quad \tilde{x} = \frac{3}{4\pi} \frac{2m_{\text{PBH}}}{a_{\text{eq}}^3 \rho_{\text{eq}}}, \quad (7.9)$$

a_{eq} and ρ_{eq} are the scale factor and total energy density at matter-radiation equality. The merger rate at time t_r is then given by

$$\begin{aligned} R(t_r) &= \int_0^{\tilde{x}} dx \int_x^\infty dy \frac{\partial^2 n_3}{\partial x \partial y} \delta(t_r - \tau(x, y)) \\ &= \frac{9 \tilde{N}_{\text{PBH}}^{53/37}}{296\pi \delta_{\text{dc}} \tilde{x}^3 \tilde{\tau}} \left(\frac{t_r}{\tilde{\tau}} \right)^{-34/37} \\ &\times \left(\Gamma \left[\frac{58}{37}, \tilde{N}_{\text{PBH}} \left(\frac{t_r}{\tilde{\tau}} \right)^{3/16} \right] - \Gamma \left[\frac{58}{37}, \tilde{N}_{\text{PBH}} \left(\frac{t_r}{\tilde{\tau}} \right)^{-1/7} \right] \right) \end{aligned} \quad (7.10)$$

with the incomplete gamma function Γ and $\tilde{N}_{\text{PBH}} = N_{\text{PBH}}(\tilde{x})$. By substituting $t_r \rightarrow t_r + \tau_{f_r}$ we can therefore compute the rate $R(t_r + \tau_{f_r})$ for the emission of this frequency f_r . Multiple merger steps, possibly present due to successive hierarchical merging of PBH binaries with increasing mass due to large clustering, can be easily implemented by adding the rates and contributions to the GW energy density parameter for the corresponding steps as detailed in ref. [334]. We include multiple merger steps when discussing the case of significant clustering, noting that this slightly shifts our results for $\delta_{\text{dc}} = 10^3$ to lower PBH abundances compared to only considering a single step.

To illustrate the importance of the time when a given frequency is emitted we show in the left panel of fig. 7.1 the rate for binaries merging at t_r (dotted blue line) with the one for GW emission with frequencies $f = 1 \mu\text{Hz}$ and 2.5 nHz today (solid green and orange lines), assuming $m_{\text{PBH}} = 10^5 M_\odot$, $f_{\text{PBH}} = 1$, and $\delta_{\text{dc}} = 1$ in each case. For instance, at $t_r = 10^8 \text{ yr}$ (i.e. $z \approx 30$) one obtains $\tau_{f_r}((1+z)1 \mu\text{Hz}) \approx 130 \text{ yr}$ and $\tau_{f_r}((1+z)2.5 \text{ nHz}) \approx 1.1 \times 10^9 \text{ yr}$. Hence, the rate for the emission of GWs with the larger frequency (solid light green line) is very close to the merger rate (dotted line), whereas the rate for the emission of GWs with the smaller frequency (solid dark green line) differs significantly, i.e. it takes the value that the dotted blue line attains $1.1 \times 10^9 \text{ yr}$ later.

The correction $t_r \rightarrow t_r + \tau_{f_r}$ is important!

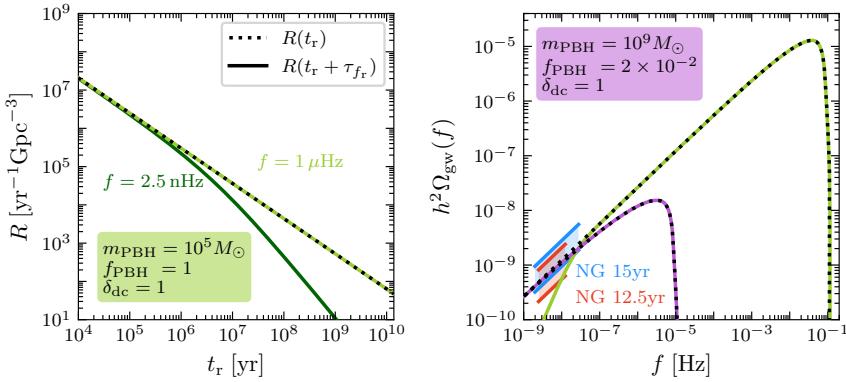


Figure 7.1: *Left:* Rate for binaries merging at t_r (dotted line) as well as rate for emitting GWs at t_r with frequencies today $f = 1 \mu\text{Hz}$ and 2.5 nHz (solid light green and dark green lines) assuming $m_{\text{PBH}} = 10^5 M_\odot$, $f_{\text{PBH}} = 1$, and $\delta_{\text{dc}} = 1$. *Right:* GW energy density $h^2 \Omega_{\text{gw}}(f)$ obtained using the correct rate $R(t_r + \tau_{f_r})$ (solid lines) and instead using the merger rate $R(t_r)$ in eq. (7.2) (dotted lines) for $m_{\text{PBH}} = 10^5 M_\odot$, $f_{\text{PBH}} = 1$, and $\delta_{\text{dc}} = 1$ (green) and $m_{\text{PBH}} = 10^9 M_\odot$, $f_{\text{PBH}} = 2 \times 10^{-2}$, and $\delta_{\text{dc}} = 1$ (purple). We also show the region where the NANOGrav 12.5yr and 15yr signals are located assuming a power-law with slope $\gamma = 13/3$ (see eq. (3.35)).

In the right panel of fig. 7.1 we show $h^2 \Omega_{\text{gw}}(f)$ as a function of the GW frequency today according to eq. (7.2) (solid lines) as well as just inserting the merger rate $R(t_r)$ instead of $R(t_r + \tau_{f_r})$ in the integral (dotted lines). The difference between those calculations is especially important for the low frequencies observed by NANOGrav if the PBH mass is relatively light.

We close the discussion of the GWB signal by mentioning some assumptions that entered in its calculation. These include

- a monochromatic mass distribution for the PBHs [336],
- a circular orbit of the PBH binaries for the GWB spectrum and the time until coalescence for a given frequency,
- neglecting the effect of other PBHs on the binaries [341, 342] as well as other environmental effects e.g. due to accretion, and
- late-time formation of binaries [336].

Assumptions

Apart from the effect of other PBHs which can potentially significantly reduce their merger rate on binaries [342], these assumptions are not expected to have a qualitative impact on our results. A valuable test of these assumption is within reach using astrophysical N -body simulations, which are, however, computationally expensive and hence beyond the scope of this thesis.

7.3 EXPECTED NUMBER OF BINARIES

On the stochasticity of Ω_{gw}

The PTA signals are reported as stochastic GWBs. While there have been searches for signals of individual binaries in the data of different PTAs [15, 264, 343], no compelling evidence for these signals was found. For sufficiently large PBH masses and small abundances the expected signal will in general no longer resemble a stochastic background, as only very few binaries will contribute to the signal, causing inconsistency with observations.

Good predictions require large \bar{N}

The problem is exacerbated by the question of how well an actual distribution of merging binaries observed by a PTA (corresponding to a local² value of GW energy density $\Omega_{\text{gw,loc}}$) would reproduce the global average of the GW energy density $h^2 \Omega_{\text{gw}}(f) = h^2 \langle \Omega_{\text{gw,loc}}(f) \rangle$, i.e. how well the global mean is reproduced by the binaries that are in our past light cone. As shown in ref. [272], even relatively generic distributions of binary mergers, scaling with the luminosity distance squared, can lead to wide distributions in $\Omega_{\text{gw,loc}}$, with width-to-mean ratios scaling like $\Delta \Omega_{\text{gw,loc}} / \Omega_{\text{gw}} \propto \bar{N}^{-1/3}$ with \bar{N} the expected number of binaries contributing to a certain frequency range, slower than predicted from the central limit theorem $\propto \bar{N}^{-1/2}$.

On calculating \bar{N}

To calculate the average number of binaries contributing to a certain frequency range, recall that $R(t_r + \tau_{f_r})$ is the number of binaries per comoving volume V_c and per time interval in the cosmic rest frame t_r emitting with a frequency f_r . Hence, the number of binaries $d\bar{N}$ emitting at redshifts between z and $z+dz$ with a frequency within the logarithmic interval $d \ln f_r$ around f_r is [344]

$$\begin{aligned} R(t_r + \tau_{f_r}) &= \frac{d^2 \bar{N}}{dt_r dV_c} = \frac{dz}{dt_r} \frac{d^2 \bar{N}}{dz d \ln f_r} \frac{d \ln f_r}{dt_r} \frac{dt_r}{dz} \frac{dz}{dV_c} \\ &= \frac{d^2 \bar{N}}{dz d \ln f_r} \left(-\frac{d \ln f_r}{d \tau_{f_r}} \right) \frac{dz}{dV_c}, \end{aligned} \quad (7.11)$$

where we used that the time in the cosmic rest frame a source is emitting within the frequency interval is given by

$$dt_r \frac{d \ln f_r}{dt_r} = -dt_r \frac{d \ln f_r}{d \tau_{f_r}}, \quad (7.12)$$

as τ_{f_r} is the time until coalescence for f_r . Since the binaries emitting in the logarithmic frequency interval $d \ln f_r$ around f_r are detected today in a logarithmic frequency interval $d \ln f$ around f , where $d \ln f_r = df_r/f_r = df/f = d \ln f$, we have

$$\frac{d^2 \bar{N}}{dz d \ln f} = \frac{d^2 \bar{N}}{dz d \ln f_r}. \quad (7.13)$$

² Local on the scales relevant for a GW signal in PTA observations, i.e. at least within a sphere around earth with a few light-years radius such that correlations between different pulsars can be affected.

With eq. (7.1) and the definition of the comoving volume in a spatially flat universe

$$V_c(z) = \frac{4\pi}{3}[d_c(z)]^3 = \frac{4\pi}{3} \left(\int_0^z \frac{dz'}{H(z')} \right)^3, \quad (7.14)$$

where d_c is the comoving distance and H is the Hubble rate, we find

$$\frac{d^2\bar{N}}{dz d \ln f} = \frac{8}{3} \tau_{f_r} \frac{4\pi [d_c(z)]^2}{H(z)} R(t_r(z) - \tau_{f_r}). \quad (7.15)$$

The average number of binaries $\bar{N}(f_-, f_+)$ contributing to GWs of frequencies between f_- and f_+ is therefore given by³

$$\bar{N}(f_-, f_+) = \int_{f_-}^{f_+} \frac{df}{f} \int_0^\infty dz \frac{8}{3} \tau_{f_r} \frac{4\pi [d_c(z)]^2}{H(z)} R(t_r(z) - \tau_{f_r}). \quad (7.16)$$

As we are interested in the number of binaries contributing to the NANOGrav 15yr signal, we take the frequency interval to span over the corresponding lowest 14 Fourier modes, i.e. $f_- = 1/T_{\text{obs}}^{15\text{yr}} = 1.98 \text{ nHz}$ and $f_+ = 14/T_{\text{obs}}^{15\text{yr}} = 27.7 \text{ nHz}$, cf. fig. 2.8.

The signal prediction for PTAs is crucially affected by the expected number of binaries emitting in the relevant frequency band [272]. Generally, the local GWB $h^2\Omega_{\text{gw,loc}}(f)$ that may be observed by a PTA is obtained by adding individual PBH binaries (the total number drawn from a Poisson distribution with mean $\bar{N}(f_-, f_+)$) distributed according to $d^2\bar{N}/(dz d \ln f)$. In particular, $h^2\Omega_{\text{gw,loc}}(f)$ is *not* deterministic given model parameters (mass, abundance, and clustering of PBHs), but instead stochastic in itself. The statistics of $h^2\Omega_{\text{gw,loc}}(f)$ can be evaluated using Markov chain Monte Carlo methods or moment generating functions [272]. Due to the considerable additional numerical effort we leave this for future studies, use the global average signal prediction $h^2\Omega_{\text{gw}}(f)$, and identify regions in parameter space where we expect a significant deviation from the global average. If $\bar{N}(f_-, f_+) \gg 1$, a lot of different PBH binaries contribute to the GW signal and the local GWB $h^2\Omega_{\text{gw,loc}}(f)$ is close to the global average $h^2\Omega_{\text{gw}}(f)$. This holds even though uncertainties due to some close-by binaries can be relevant for $\bar{N}(f_-, f_+) \lesssim 100$ [272]. If $\bar{N}(f_-, f_+) \sim \mathcal{O}(\text{few})$, i.e. if the GW signal is composed of only a few binaries, the uncertainty in the signal prediction is considerable. Arguably, it would be more appropriate to search for individual GW events instead of a GWB [15, 264, 343]. If $\bar{N}(f_-, f_+) \ll 1$, even having a single PBH binary emitting GWs is unlikely and, though it is possible to have an unexpectedly large background as there is a single binary nonetheless, in most realizations $h^2\Omega_{\text{gw,loc}}(f) = 0$.

High signal prediction uncertainty for $\bar{N} \sim \mathcal{O}(\text{few})$

³ The *actual* number of binaries is then Poisson distributed with mean $\bar{N}(f_-, f_+)$.

7.4 PBH PRODUCTION AND CONSTRAINTS

We remain agnostic about the PBH production

There are many different production mechanisms for PBHs, see e.g. [35, 254, 255, 345–348]. Highly clustered PBH distributions in particular are not expected for Gaussian primordial fluctuations [349, 350], but could e.g. arise due to primordial non-Gaussianities [332, 351, 352] or from the collapse of domain walls [346]. In this thesis we remain agnostic about the origin and spatial distribution of PBHs and concentrate on exploring the phenomenological impact of different assumptions.

Different astrophysical and cosmological observations place constraints on the abundance of heavy PBHs, which we adopt from [28]. These limits assume a monochromatic mass function as well as a roughly homogeneous spatial distribution (no clustering). While we briefly comment on the expected impact of sizable clustering, a full re-evaluation of these limits would require going beyond some of the simple approximations made in the original derivations and is beyond the scope of this thesis. Also note that many of the different constraints come with different uncertainties and sometimes also with additional caveats.

The most relevant PBH constraints

The most relevant limits in the mass range of interest come from the heating of stars in the Galactic disk [353], the tidal disruption of galaxies [353], the dynamical friction effect on halo objects [353], requiring successful formation of the observed large scale structure [34], and observations of X-ray binaries [354]. Many of these limits require at least one PBH per relevant cosmic structure. In case of PBH clustering, we expect that some structures will contain more than one PBH while others contain no PBH at all. This will likely weaken a number of limits and move them to smaller masses. A proper evaluation of the PBH spatial distribution and the resulting limits will however require detailed simulations.

The case of μ -distortions

Depending on the production mechanism of the PBHs, strong limits may also arise from the observation of the CMB. In particular if PBHs form due to the tail of Gaussian density fluctuations, Silk damping leads to μ -distortions and strong constraints over a sizable mass range [355]. However, this limit crucially depends on the production mechanism and may therefore even be completely evaded, see for instance ref. [332]. In fact, as discussed above, large clustering generally requires a different production mechanism, calling into question the relevance of these limits.

7.5 PTA DATA ANALYSIS

We fit the GWB spectrum from PBH binary mergers to the NANOGrav 15yr and 12.5yr data sets via the interface `ptarcade` [356] for `ceffyl` [118] using the 14 (five) lowest Fourier modes for the 15yr (12.5yr) data set. Given that evidence for an HD correlation is only present in the new

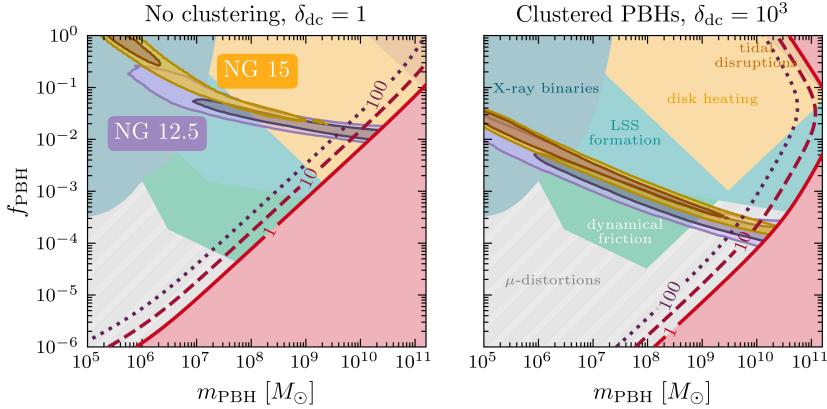


Figure 7.2: Best-fit regions where the NANOGrav 15yr (orange) and 12.5yr (purple) data set can be explained by inspiraling PBH binaries without clustering (left) and with significant clustering (right). We indicate regions, where $\bar{N} < 1$, 10, and 100 PBH binaries are expected to contribute to the signal, and show complementary constraints as discussed in the main text.

data set, we assume a CURN spectrum for the 12.5yr data set and only move to HD correlations for the 15yr data set. We validated our approach by comparison to results obtained using `enterprise` and `enterprise-extensions` [115, 116]. We perform calculations with no clustering, $\delta_{dc} = 1$, and significant clustering, $\delta_{dc} = 10^3$, choosing log priors $f_{\text{PBH}} \in [10^{-5}, 1]$ and $m_{\text{PBH}} \in [10^5, 10^{12}] M_\odot$.

To obtain constraints on the (clustered) PBH scenario, we perform an additional scan over an extended model parameter space $m_{\text{PBH}}, f_{\text{PBH}}, A_{\text{GWB}}$ including an additional mock GW signal contribution (with $\gamma = 13/3$ and $A_{\text{GWB}} \in [10^{-18}, 10^{-14}]$, cf. eq. (2) in ref. [238]) and then marginalize over A_{GWB} . The region in the remaining $(m_{\text{PBH}}, f_{\text{PBH}})$ plane excluded with 2σ corresponds to the parameter space in which the GW signal would be too strong to account for the signal. To obtain our constraints on clustered PBHs we further conservatively remove the parameter space in which less than $\bar{N} = 10$ merger events would contribute to the NANOGrav signal.

7.6 RESULTS

In fig. 7.2 we show the regions in PBH mass m_{PBH} and DM fraction f_{PBH} , where the NANOGrav signal can be explained assuming merging PBHs with no clustering (left) and significant clustering with $\delta_{dc} = 10^3$ (right). These contours are only expected to be reliable when the expected number of binaries is $\bar{N} \gtrsim 10$ (cf. dashed dark red lines) as otherwise the signal observable in NANOGrav is expected to have sig-

Our analysis pipeline

Deriving PBH constraints

Homogeneously distributed PBHs cannot explain the signal

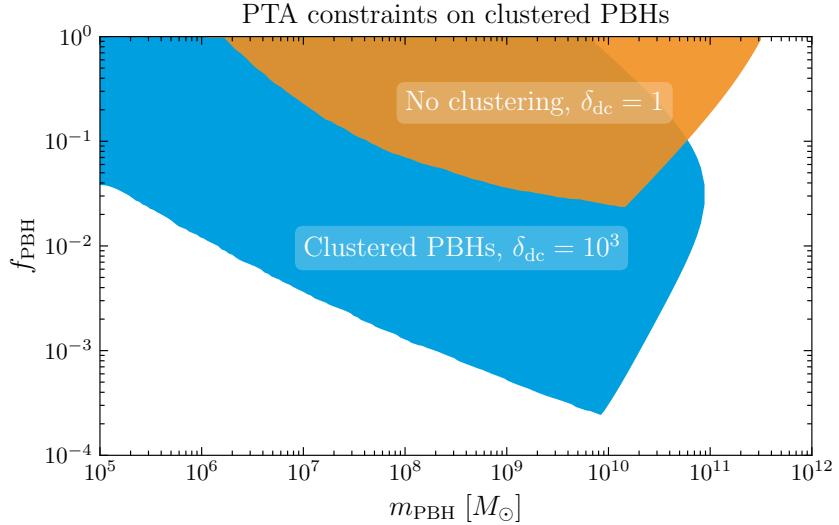


Figure 7.3: Constraints on PBHs from PTA observations with no clustering (orange) and significant clustering $\delta_{dc} = 10^3$ (blue). We conservatively cut the limits when the expected number of binaries contributing to the signal falls below $\bar{N} = 10$.

nificant deviations from the global average GWB signal used in our analysis. Especially for $\bar{N} < 1$ (full red line) one would not even expect any GWB signal in most realizations of PBH binary distributions. These effects were neglected in [333] that falsely concluded that the NANOGrav signal could be explained in a region of parameter space without clustering, where $\bar{N} \lesssim 1$. We find that the case without clustering cannot explain the NANOGrav signal once taking into account cosmological and astrophysical constraints.

Including clustering shifts the signal regions to smaller f_{PBH} , enabling a consistent explanation of the PTA data without violating observational constraints, provided that the PTA formation mechanism does not result in significant μ -distortions. Note that clustering is also expected to further open up parameter space for a consistent explanation as complementary constraints are expected to become weaker.

Comparing the results for the 15yr and 12.5yr data sets we note that the signal regions have shifted to larger f_{PBH} , due to the preference for a larger GWB in the new data set [7, 238], and there is a slight preference for lower masses, where the earlier emission of GWs leads to an increased slope, cf. fig. 7.1, as preferred by the new data [7].

If the GWB signal generated by inspiraling PBH binaries becomes too large, above the regions explaining the PTA data in fig. 7.2, the abundance of PBHs can be *constrained* by PTA observations. The constraints are shown in fig. 7.3 with and without clustering, conservatively requiring $\bar{N} \geq 10$.

*Clustering saves
the explanation*

*15yr vs. 12.5yr
data*

*Novel PTA
constraints*

7.7 DISCUSSION AND CONCLUSIONS

In this chapter we have studied the possibility that the signal observed by various PTAs is due to inspiraling primordial SMBHBs. If the PBHs are “homogeneously” distributed at their formation, i.e. follow a Poisson distribution, significant cosmological and astrophysical constraints exclude the possibility of explaining the PTA signal with inspiraling PBHs. Instead considering a clustered spatial distribution of PBHs increases the binary merger rate and thus enables a consistent explanation of the PTA signals with merging PBH binaries. Crucially, we have checked that also the signal prediction is reliable in the relevant parameter space by computing the expected number of binaries contributing to the GW signal. Further, we used PTA data to constrain the PBH parameter space when the GWB generated during the inspiral would result in stronger signal strengths than the one detected.

The studied scenario may serve as a motivation for model builders to construct scenarios that can generate clustered supermassive PBHs without running into cosmological and astrophysical constraints, in particular due to μ -distortions of the CMB arising from PBH formation. While the latest PTA data finds no evidence for individual compact binary merger events on top of a stochastic background or anisotropies of the GW spectrum, the situation might change in the future. Such a detection would likely invalidate most other cosmological explanations, while being a prediction of our scenario.

PTAs could have detected inspiral of clustered PBH

Our model also features anisotropies

8 | SUMMARY

The wild ideas of yesterday quickly become today's dogma.

— Sheldon Glashow, Physics Nobel laureate 1979

In this thesis, we have explored the dawn of GW cosmology, an emerging field that promises to revolutionize our understanding of the early universe. Traditionally, BBN has served as the earliest probe of the Λ CDM model, allowing us to make precise statements up to temperatures around the MeV scale. However, the recent detection of a gravitational wave background (GWB) by pulsar timing arrays (PTAs) has provided us with the first direct probe of much higher temperatures, reaching up to the GeV scale.

Specifically, we have demonstrated in chapter 5 that dark sector phase transitions (DSPTs) occurring at around $T_p \simeq 10$ MeV could explain the observed GWB. In a first step we found that a phase transition within a secluded dark sector is in significant tension with precision observables from BBN and the CMB: Ignoring constraints on the effective number of neutrino species (N_{eff}), one could incorrectly conclude that arbitrarily strong DSPTs could explain the PTA signal (see fig. 5.1). However, when including these constraints, we find that only transitions with a strength parameter $\alpha < 0.1$ and extremely small $\beta/H \simeq \mathcal{O}(1-10)$, corresponding to very slow transitions, are favored (see fig. 5.3). Requiring $\beta/H > 3$ in order to have bubbles of sub-Hubble size emitting the GW signal yields a reasonably good fit to the data. However, further requiring $\beta/H > 10$ to avoid overestimating the GWB, leads to the conclusion that no good match to the PTA data can be obtained when cosmological constraints are included within our global fit framework.

Thereafter, we have shown that the fit to the data can be significantly improved, and the DSPT interpretation preserved, if the dark sector is not secluded but instead decays into the photon bath before the onset of BBN. Modeling this through the decay of a 5 MeV scalar into photons in our global fit, we found that the maximal allowed lifetime of the dark Higgs boson is $\tau_\phi = \mathcal{O}(0.1)$ s, which is below the sensitivity of laboratory experiments (see fig. 5.5). The central conclusions of this chapter are anticipated to remain robust even when updated with the new 15yr NANOGrav data set, replacing the 12.5yr data set that was the most recent available at the time our analysis was conducted.

In the study presented in chapter 6, we examined the implications of a future detection of a GWB with LISA when interpreted as a DSPT. Specifically, we considered the spontaneous breaking of a $U(1)'$ symmetry, triggering the freeze-out of a fermionic dark matter candidate.

The meaning of the PTA results for cosmology

Cosmological constraints challenge stable DSPTs

PTAs could have observed a decaying DSPT

*A FOPT-triggered
freeze-out emits
GWs in LISA band*

*The correlation is
not spoiled for hot
DSs, allowing
stronger GWBs*

*Inspiraling PBHs
can explain
NANOGrav*

*This thesis was
only the start*

The road ahead

We uncovered a previously underappreciated correlation between the peak frequency of the produced GW signal and the relic abundance of dark matter (see fig. 6.5). Fixing the produced DM abundance to the observed relic density leads to the prediction of a GWB within the LISA band, see fig. 6.6. We performed extensive scans over the $U(1)'$ model parameter space and showed that this conclusion remains a robust feature which could potentially also be present in other dark sector models.

Further, we investigated in detail how the dark sector transfers energy to the SM bath and whether it is justified to assume a common temperature for both sectors. We found that the two sectors quickly thermalize for portal couplings of $\lambda_{h\phi} \gtrsim 10^{-6}$ between the dark Higgs and the visible Higgs, a value small enough to satisfy laboratory constraints (see figs. 6.9). Even when treating $\lambda_{h\phi} < 10^{-6}$ and the initial temperature ratio between the two sectors (the latter having a strong impact on the amplitude of the emitted GWB) as open parameters, the correlation between the GWB's peak frequency and the relic abundance remains robust (see fig. 6.12).

In our final chapter 7, we explored the conditions under which the inspiral of PBHs could account for the GWB observed by PTAs. We found that while the signal can be explained by inspiraling PBHs, this is only possible if the PBHs are initially clustered, contrary to previous claims (see fig. 7.2). We argue that there exists a part of parameter space in which the signal can be explained while avoiding astrophysical constraints, provided that the PBHs are formed circumventing μ -distortion limits. This explanation of the PTA signal is of particular relevance, as it would not be ruled out in case future PTA observations find signs of strong anisotropy in the GWB spectrum, which indicate a tension with early universe explanations like phase transitions.

In summary, this thesis has contributed to the burgeoning field of GW cosmology. It highlights the crucial role that BBN and CMB observations will play in interpreting PTA results and performing model comparisons in global fit frameworks. Looking forward, future observatories like LISA will expand our observational capabilities to phenomena occurring as early as picoseconds after the end of inflation, possibly helping us in unraveling the origin of dark matter in our universe.

The interplay between GW science and dark matter phenomenology is still in its early stages, with much more to explore. This thesis lays the groundwork for understanding the intimate relationship between violent early universe processes like phase transitions or the inspiral of supermassive primordial black holes and their respective GW emission. The key takeaway is clear: gravitational wave backgrounds can provide a unique window to the *dark universe* and offer new insights into the fundamental processes that shaped our cosmos.

BIBLIOGRAPHY

- [1] T. Bringmann, P. F. Depta, T. Konstandin, K. Schmidt-Hoberg, and C. Tasillo, *Does NANOGrav observe a dark sector phase transition?*, *JCAP* **11** (2023) 053, [[2306.09411](#)].
- [2] T. Bringmann, T. E. Gonzalo, F. Kahlhoefer, J. Matuszak, and C. Tasillo, *Hunting WIMPs with LISA: correlating dark matter and gravitational wave signals*, *JCAP* **05** (2024) 065, [[2311.06346](#)].
- [3] P. F. Depta, K. Schmidt-Hoberg, P. Schwaller, and C. Tasillo, *Do pulsar timing arrays observe merging primordial black holes?*, [[2306.17836](#)].
- [4] F. Ertas, F. Kahlhoefer, and C. Tasillo, *Turn up the volume: listening to phase transitions in hot dark sectors*, *JCAP* **02** (2022), no. 02 014, [[2109.06208](#)].
- [5] **LIGO Scientific, Virgo Collaboration**, B. P. Abbott et al., *Observation of Gravitational Waves from a Binary Black Hole Merger*, *Phys. Rev. Lett.* **116** (2016), no. 6 061102, [[1602.03837](#)].
- [6] A. Einstein, *Approximative Integration of the Field Equations of Gravitation*, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1916** (1916) 688–696.
- [7] **NANOGrav Collaboration**, G. Agazie et al., *The NANOGrav 15-year Data Set: Evidence for a Gravitational-Wave Background*, *Astrophys. J. Lett.* **951** (2023), no. 1 [[2306.16213](#)].
- [8] D. J. Reardon et al., *Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array*, *Astrophys. J. Lett.* **951** (2023), no. 1 L6, [[2306.16215](#)].
- [9] **EPTA Collaboration**, J. Antoniadis et al., *The second data release from the European Pulsar Timing Array III. Search for gravitational wave signals*, *Astron. Astrophys.* **678** (2023) A50, [[2306.16214](#)].
- [10] H. Xu et al., *Searching for the Nano-Hertz Stochastic Gravitational Wave Background with the Chinese Pulsar Timing Array Data Release I*, *Res. Astron. Astrophys.* **23** (2023), no. 7 075024, [[2306.16216](#)].
- [11] C. Caprini and D. G. Figueroa, *Cosmological Backgrounds of Gravitational Waves*, *Class. Quant. Grav.* **35** (2018), no. 16 163001, [[1801.04268](#)].

- [12] **Planck Collaboration**, N. Aghanim et al., *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6, [[1807.06209](#)]. [Erratum: *Astron.Astrophys.* 652, C4 (2021)].
- [13] D. Baumann, *Cosmology*. Cambridge University Press, 7, 2022.
- [14] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford University Press, 2007.
- [15] **NANOGrav Collaboration**, G. Agazie et al., *The NANOGrav 15 yr Data Set: Bayesian Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries*, *Astrophys. J. Lett.* **951** (2023), no. 2 L50, [[2306.16222](#)].
- [16] J. Binney and S. Tremaine, *Galactic dynamics*. Princeton University Press, 1987.
- [17] P. B. Ivanov, J. C. B. Papaloizou, and A. G. Polnarev, *The evolution of a supermassive binary caused by an accretion disc*, *Mon. Not. Roy. Astron. Soc.* **307** (1999) 79, [[astro-ph/9812198](#)].
- [18] Z. Haiman, Z. Haiman, B. Kocsis, B. Kocsis, K. Menou, et al., *The Population of Viscosity- and Gravitational Wave-Driven Supermassive Black Hole Binaries Among Luminous AGN*, *Astrophys. J.* **700** (2009) 1952–1969, [[0904.1383](#)]. [Erratum: *Astrophys.J.* 937, 129 (2022)].
- [19] P. Amaro-Seoane, A. Sesana, L. Hoffman, M. Benacquista, C. Eichhorn, et al., *Triplets of supermassive black holes: Astrophysics, Gravitational Waves and Detection*, *Mon. Not. Roy. Astron. Soc.* **402** (2010) 2308, [[0910.1587](#)].
- [20] B. Kocsis and A. Sesana, *Gas driven massive black hole binaries: signatures in the nHz gravitational wave background*, *Mon. Not. Roy. Astron. Soc.* **411** (2011) 1467, [[1002.0584](#)].
- [21] E. Vasiliev and D. Merritt, *The loss cone problem in axisymmetric nuclei*, *Astrophys. J.* **774** (2013) 87, [[1301.3150](#)].
- [22] E. Vasiliev, F. Antonini, and D. Merritt, *The final-parsec problem in nonspherical galaxies revisited*, *Astrophys. J.* **785** (2014) 163, [[1311.1167](#)].
- [23] E. Vasiliev, F. Antonini, and D. Merritt, *The Final-parsec Problem in the Collisionless Limit*, *APJ* **810** (Sept., 2015) 49, [[1505.05480](#)].
- [24] F. Dosopoulou and F. Antonini, *Dynamical friction and the evolution of Supermassive Black hole Binaries: the final hundred-parsec problem*, *Astrophys. J.* **840** (2017), no. 1 31, [[1611.06573](#)].
- [25] M. Breitbach, J. Kopp, E. Madge, T. Opferkuch, and P. Schwaller, *Dark, Cold, and Noisy: Constraining Secluded Hidden Sectors*

- with Gravitational Waves*, *JCAP* **07** (2019) 007, [[1811.11175](#)].
- [26] P. Schwaller, *Gravitational Waves from a Dark Phase Transition*, *Phys. Rev. Lett.* **115** (2015), no. 18 181101, [[1504.07263](#)].
- [27] C. Caprini et al., *Science with the space-based interferometer eLISA. II: Gravitational waves from cosmological phase transitions*, *JCAP* **04** (2016) 001, [[1512.06239](#)].
- [28] B. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, *Constraints on primordial black holes*, *Rept. Prog. Phys.* **84** (2021), no. 11 116902, [[2002.12778](#)].
- [29] M. Volonteri, *Formation of Supermassive Black Holes*, *Astron. Astrophys. Rev.* **18** (2010) 279–315, [[1003.4404](#)].
- [30] M. Volonteri, M. Habouzit, and M. Colpi, *The origins of massive black holes*, *Nature Rev. Phys.* **3** (2021), no. 11 732–743, [[2110.10175](#)].
- [31] S. L. Shapiro, *Spin, accretion and the cosmological growth of supermassive black holes*, *Astrophys. J.* **620** (2005) 59–68, [[astro-ph/0411156](#)].
- [32] M. Volonteri and M. J. Rees, *Quasars at z=6: The survival of the fittest*, *Astrophys. J.* **650** (2006) 669–678, [[astro-ph/0607093](#)].
- [33] T. Tanaka and Z. Haiman, *The Assembly of Supermassive Black Holes at High Redshifts*, *Astrophys. J.* **696** (2009) 1798–1822, [[0807.4702](#)].
- [34] B. Carr and J. Silk, *Primordial Black Holes as Generators of Cosmic Structures*, *Mon. Not. Roy. Astron. Soc.* **478** (2018), no. 3 3756–3775, [[1801.00672](#)].
- [35] Y. Gouttenoire and T. Volansky, *Primordial Black Holes from Supercooled Phase Transitions*, [[2305.04942](#)].
- [36] Y. Gouttenoire and E. Vitagliano, *Primordial black holes and wormholes from domain wall networks*, *Phys. Rev. D* **109** (2024), no. 12 123507, [[2311.07670](#)].
- [37] M. Maggiore, *Gravitational Waves. Vol. 2: Astrophysics and Cosmology*. Oxford University Press, 3, 2018.
- [38] E. W. Kolb and M. S. Turner, *The Early Universe*, vol. 69. Addison-Wesley, 1990.
- [39] **Particle Data Group Collaboration**, R. L. Workman et al., *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.
- [40] F. Gao, S. Sun, and G. White, *A first-order deconfinement phase transition in the early universe and gravitational waves*, [[2405.00490](#)].

- [41] J. Lesgourges, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, [[1104.2932](#)].
- [42] N. Schöneberg, G. Franco Abellán, A. Pérez Sánchez, S. J. Witte, V. Poulin, et al., *The H0 Olympics: A fair ranking of proposed models*, *Phys. Rept.* **984** (2022) 1–55, [[2107.10291](#)].
- [43] W. L. Freedman and B. F. Madore, *Progress in direct measurements of the Hubble constant*, *JCAP* **11** (2023) 050, [[2309.05618](#)].
- [44] W. L. Freedman, B. F. Madore, I. S. Jang, T. J. Hoyt, A. J. Lee, et al., *Status Report on the Chicago-Carnegie Hubble Program (CCHP): Three Independent Astrophysical Determinations of the Hubble Constant Using the James Webb Space Telescope*, [[2408.06153](#)].
- [45] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, *Helv. Phys. Acta* **6** (1933) 110–127.
- [46] V. C. Rubin and W. K. Ford, Jr., *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, *Astrophys. J.* **159** (1970) 379–403.
- [47] L. van Waerbeke et al., *Detection of correlated galaxy ellipticities on CFHT data: First evidence for gravitational lensing by large scale structures*, *Astron. Astrophys.* **358** (2000) 30–44, [[astro-ph/0002500](#)].
- [48] D. J. Bacon, A. R. Refregier, and R. S. Ellis, *Detection of weak gravitational lensing by large-scale structure*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 625, [[astro-ph/0003008](#)].
- [49] D. M. Wittman, J. A. Tyson, D. Kirkman, I. Dell’Antonio, and G. Bernstein, *Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales*, *Nature* **405** (2000) 143–149, [[astro-ph/0003014](#)].
- [50] R. Massey, T. Kitching, and J. Richard, *The dark matter of gravitational lensing*, *Rept. Prog. Phys.* **73** (2010) 086901, [[1001.1739](#)].
- [51] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, et al., *A direct empirical proof of the existence of dark matter*, *Astrophys. J. Lett.* **648** (2006) L109–L113, [[astro-ph/0608407](#)].
- [52] D. Harvey, R. Massey, T. Kitching, A. Taylor, and E. Tittley, *The non-gravitational interactions of dark matter in colliding galaxy clusters*, *Science* **347** (2015) 1462–1465, [[1503.07675](#)].
- [53] A. Robertson, R. Massey, and V. Eke, *Cosmic particle colliders: simulations of self-interacting dark matter with anisotropic scattering*, *Mon. Not. Roy. Astron. Soc.* **467** (2017), no. 4

- 4719–4730, [[1612.03906](#)].
- [54] S. Profumo, *An Introduction to Particle Dark Matter*. World Scientific, 2017.
- [55] M. Cirelli, A. Strumia, and J. Zupan, *Dark Matter*, [[2406.01705](#)].
- [56] G. Arcadi, D. Cabo-Almeida, M. Dutra, P. Ghosh, M. Lindner, et al., *The Waning of the WIMP: Endgame?*, [[2403.15860](#)].
- [57] S. M. Carroll, *Spacetime and Geometry: An Introduction to General Relativity*. Cambridge University Press, 7, 2019.
- [58] R. M. Wald, *General Relativity*. Chicago Univ. Pr., Chicago, USA, 1984.
- [59] L. Husdal, *On Effective Degrees of Freedom in the Early Universe*, *Galaxies* **4** (2016), no. 4 78, [[1609.04979](#)].
- [60] J. Lesgourgues, G. Mangano, G. Miele, and S. Pastor, *Neutrino Cosmology*. Cambridge University Press, 2, 2013.
- [61] J. J. Bennett, G. Buldgen, P. F. De Salas, M. Drewes, S. Gariazzo, et al., *Towards a precision calculation of N_{eff} in the Standard Model II: Neutrino decoupling in the presence of flavour oscillations and finite-temperature QED*, *JCAP* **04** (2021) 073, [[2012.02726](#)].
- [62] S. Brieden, H. Gil-Marín, and L. Verde, *Model-agnostic interpretation of 10 billion years of cosmic evolution traced by BOSS and eBOSS data*, *JCAP* **08** (2022), no. 08 024, [[2204.11868](#)].
- [63] J. McDonald, *WIMP Densities in Decaying Particle Dominated Cosmology*, *Phys. Rev. D* **43** (1991) 1063–1068.
- [64] L. J. Hall, K. Jedamzik, J. March-Russell, and S. M. West, *Freeze-In Production of FIMP Dark Matter*, *JHEP* **03** (2010) 080, [[0911.1120](#)].
- [65] K. Griest and D. Seckel, *Three exceptions in the calculation of relic abundances*, *Phys. Rev. D* **43** (1991) 3191–3203.
- [66] R. T. D’Agnolo and J. T. Ruderman, *Light Dark Matter from Forbidden Channels*, *Phys. Rev. Lett.* **115** (2015), no. 6 061301, [[1505.07107](#)].
- [67] A. D. Dolgov, *ON CONCENTRATION OF RELICT THETA PARTICLES. (IN RUSSIAN)*, *Yad. Fiz.* **31** (1980) 1522–1528.
- [68] E. D. Carlson, M. E. Machacek, and L. J. Hall, *Self-interacting dark matter*, *Astrophys. J.* **398** (1992) 43–52.
- [69] T. Bringmann, P. F. Depta, M. Hufnagel, J. T. Ruderman, and K. Schmidt-Hoberg, *Dark Matter from Exponential Growth*,

- Phys. Rev. Lett.* **127** (2021), no. 19 191802, [2103.16572].
- [70] M. Cirelli, Y. Gouttenoire, K. Petraki, and F. Sala, *Homeopathic Dark Matter, or how diluted heavy substances produce high energy cosmic rays*, *JCAP* **02** (2019) 014, [1811.03608].
- [71] M. Kamionkowski and M. S. Turner, *THERMAL RELICS: DO WE KNOW THEIR ABUNDANCES?*, *Phys. Rev. D* **42** (1990) 3310–3320.
- [72] C. Pitrou, A. Coc, J.-P. Uzan, and E. Vangioni, *Precision Big Bang Nucleosynthesis with the New Code PRIMAT*, *JPS Conf. Proc.* **31** (2020) 011034, [1909.12046].
- [73] A. Arbey, *AlterBBN: A program for calculating the BBN abundances of the elements in alternative cosmologies*, *Comput. Phys. Commun.* **183** (2012) 1822–1831, [1106.1363].
- [74] T.-H. Yeh, J. Shelton, K. A. Olive, and B. D. Fields, *Probing physics beyond the standard model: limits from BBN and the CMB independently and combined*, *JCAP* **10** (2022) 046, [2207.13133].
- [75] A. Einstein, *Über Gravitationswellen, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1918** (1918) 154–167.
- [76] A. S. Eddington, *The propagation of gravitational waves*, *Proc. Roy. Soc. Lond. A* **102** (1922) 268–282.
- [77] A. Einstein and N. Rosen, *On Gravitational waves*, *J. Franklin Inst.* **223** (1937) 43–54.
- [78] F. A. E. Pirani, *On the Physical significance of the Riemann tensor*, *Acta Phys. Polon.* **15** (1956) 389–405.
- [79] H. Bondi, *Plane gravitational waves in general relativity*, *Nature* **179** (1957) 1072–1073.
- [80] J. Weber, *Evidence for discovery of gravitational radiation*, *Phys. Rev. Lett.* **22** (1969) 1320–1324.
- [81] R. A. Hulse and J. H. Taylor, *Discovery of a pulsar in a binary system*, *Astrophys. J. Lett.* **195** (1975) L51–L53.
- [82] G. E. Moss, L. R. Miller, and R. L. Forward, *Photon-noise-limited laser transducer for gravitational antenna*, *Appl. Opt.* **10** (1971) 2495–2498.
- [83] R. Weiss, *Republication of: Electromagnetically coupled broadband gravitational antenna*, *Gen. Rel. Grav.* **54** (2022), no. 11 153.
- [84] D. Kennefick, *Traveling at the speed of thought: Einstein and the quest for gravitational waves*. Princeton, USA: Univ. Pr., 2007.

- [85] C. Uggla and J. Wainwright, *Cosmological Perturbation Theory Revisited*, *Class. Quant. Grav.* **28** (2011) 175017, [[1102.5039](#)].
- [86] A. Hook, G. Marques-Tavares, and D. Racco, *Causal gravitational waves as a probe of free streaming particles and the expansion of the Universe*, *JHEP* **02** (2021) 117, [[2010.03568](#)].
- [87] G. Franciolini, D. Racco, and F. Rompineve, *Footprints of the QCD Crossover on Cosmological Gravitational Waves at Pulsar Timing Arrays*, *Phys. Rev. Lett.* **132** (2024), no. 8 081001, [[2306.17136](#)].
- [88] A. A. Penzias and R. W. Wilson, *A Measurement of excess antenna temperature at 4080-Mc/s*, *Astrophys. J.* **142** (1965) 419–421.
- [89] D. C. N. da Cunha and C. Ringeval, *Interferences in the Stochastic Gravitational Wave Background*, *JCAP* **08** (2021) 005, [[2104.14231](#)].
- [90] K. Schmitz, *New Sensitivity Curves for Gravitational-Wave Signals from Cosmological Phase Transitions*, *JHEP* **01** (2021) 097, [[2002.04615](#)].
- [91] T. Bringmann, V. Domcke, E. Fuchs, and J. Kopp, *High-frequency gravitational wave detection via optical frequency modulation*, *Phys. Rev. D* **108** (2023), no. 6 L061303, [[2304.10579](#)].
- [92] A. Berlin, D. Blas, R. Tito D’Agnolo, S. A. R. Ellis, R. Harnik, et al., *Electromagnetic cavities as mechanical bars for gravitational waves*, *Phys. Rev. D* **108** (2023), no. 8 084058, [[2303.01518](#)].
- [93] A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott, and R. A. Collins, *Observation of a rapidly pulsating radio source*, *Nature* **217** (1968) 709–713.
- [94] L. D. Landau, *On the theory of stars*, *Phys. Z. Sowjetunion* **1** (1932) 285.
- [95] W. Baade and F. Zwicky, *On Super-novae*, *Proceedings of the National Academy of Science* **20** (May, 1934) 254–259.
- [96] F. Pacini, *Rotating Neutron Stars, Pulsars and Supernova Remnants*, *Nature* **219** (July, 1968) 145–146.
- [97] T. Gold, *Rotating neutron stars as the origin of the pulsating radio sources*, *Nature* **218** (1968) 731–732.
- [98] D. R. Lorimer and M. Kramer, *Handbook of Pulsar Astronomy*, vol. 4. Cambridge University Press, 2004.
- [99] P. Goldreich and W. H. Julian, *Pulsar electrodynamics*, *Astrophys. J.* **157** (1969) 869.

- [100] M. V. Sazhin, *Opportunities for detecting ultralong gravitational waves*, *Sov. Astron.* **22** (1978) 36–38.
- [101] S. L. Detweiler, *Pulsar timing measurements and the search for gravitational waves*, *Astrophys. J.* **234** (1979) 1100–1104.
- [102] R. w. Hellings and G. s. Downs, *UPPER LIMITS ON THE ISOTROPIC GRAVITATIONAL RADIATION BACKGROUND FROM PULSAR TIMING ANALYSIS*, *Astrophys. J. Lett.* **265** (1983) L39–L42.
- [103] R. S. Foster and D. C. Backer, *Constructing a Pulsar Timing Array*, *APJ* **361** (Sept., 1990) 300.
- [104] G. Hobbs et al., *A pulsar-based time-scale from the International Pulsar Timing Array*, *Mon. Not. Roy. Astron. Soc.* **491** (2020), no. 4 5951–5965, [[1910.13628](#)].
- [105] R. N. Caballero et al., *Studying the solar system with the International Pulsar Timing Array*, *Mon. Not. Roy. Astron. Soc.* **481** (2018), no. 4 5501–5516, [[1809.10744](#)].
- [106] **NANOGrav Collaboration**, A. Afzal et al., *The NANOGrav 15-year Data Set: Search for Signals from New Physics*, *Astrophys. J. Lett.* **951** (2023), no. 1 [[2306.16219](#)].
- [107] S. R. Taylor, *The Nanohertz Gravitational Wave Astronomer*, [[2105.13270](#)].
- [108] C. M. F. Mingarelli and T. Sidery, *Effect of small interpulsar distances in stochastic gravitational wave background searches with pulsar timing arrays*, *Phys. Rev. D* **90** (2014), no. 6 062011, [[1408.6840](#)].
- [109] R. C. Bernardo and K.-W. Ng, *Pulsar and cosmic variances of pulsar timing-array correlation measurements of the stochastic gravitational wave background*, *JCAP* **11** (2022) 046, [[2209.14834](#)].
- [110] R. C. Bernardo and K.-W. Ng, *Constraining gravitational wave propagation using pulsar timing array correlations*, *Phys. Rev. D* **107** (2023), no. 10 L101502, [[2302.11796](#)].
- [111] G. Hobbs, R. Edwards, and R. Manchester, *Tempo2, a new pulsar timing package. 1. overview*, *Mon. Not. Roy. Astron. Soc.* **369** (2006) 655–672, [[astro-ph/0603381](#)].
- [112] R. T. Edwards, G. B. Hobbs, and R. N. Manchester, *Tempo2, a new pulsar timing package. 2. The timing model and precision estimates*, *Mon. Not. Roy. Astron. Soc.* **372** (2006) 1549–1574, [[astro-ph/0607664](#)].
- [113] G. Hobbs, F. Jenet, K. J. Lee, J. P. W. Verbiest, D. Yardley, et al., *TEMPO2, a new pulsar timing package. III: Gravitational*

- wave simulation,* *Mon. Not. Roy. Astron. Soc.* **394** (2009) 1945, [[0901.0592](#)].
- [114] J. Luo et al., *PINT: A Modern Software Package for Pulsar Timing*, *Astrophys. J.* **911** (2021), no. 1 45, [[2012.00074](#)].
- [115] J. A. Ellis, M. Vallisneri, S. R. Taylor, and P. T. Baker, “Enterprise: Enhanced numerical toolbox enabling a robust pulsar inference suite.” Zenodo, Sept., 2020.
- [116] S. R. Taylor, P. T. Baker, J. S. Hazboun, J. Simon, and S. J. Vigeland, *enterprise_extensions*, 2021. v2.3.3.
- [117] J. Ellis and R. van Haasteren, *Ptmcmc sampler: Official release*, 2017.
- [118] W. G. Lamb, S. R. Taylor, and R. van Haasteren, *Rapid refitting techniques for Bayesian spectral characterization of the gravitational wave background using pulsar timing arrays*, *Phys. Rev. D* **108** (2023), no. 10 103019, [[2303.15442](#)].
- [119] Y. Nakai, M. Suzuki, F. Takahashi, and M. Yamada, *Gravitational Waves and Dark Radiation from Dark Phase Transition: Connecting NANOGrav Pulsar Timing Data and Hubble Tension*, *Phys. Lett. B* **816** (2021) 136238, [[2009.09754](#)].
- [120] W. Ratzinger and P. Schwaller, *Whispers from the dark side: Confronting light new physics with NANOGrav data*, *SciPost Phys.* **10** (2021), no. 2 047, [[2009.11875](#)].
- [121] **NANOGrav Collaboration**, Z. Arzoumanian et al., *Searching for Gravitational Waves from Cosmological Phase Transitions with the NANOGrav 12.5-Year Dataset*, *Phys. Rev. Lett.* **127** (2021), no. 25 251302, [[2104.13930](#)].
- [122] E. Madge, E. Morgante, C. P. Ibáñez, N. Ramberg, W. Ratzinger, et al., *Primordial gravitational waves in the nano-Hertz regime and PTA data – towards solving the GW inverse problem*, [[2306.14856](#)].
- [123] Y. Gouttenoire, S. Trifinopoulos, G. Valogiannis, and M. Vanvlasselaer, *Scrutinizing the primordial black hole interpretation of PTA gravitational waves and JWST early galaxies*, *Phys. Rev. D* **109** (2024), no. 12 123002, [[2307.01457](#)].
- [124] J. Ellis, M. Lewicki, C. Lin, and V. Vaskonen, *Cosmic superstrings revisited in light of NANOGrav 15-year data*, *Phys. Rev. D* **108** (2023), no. 10 103511, [[2306.17147](#)].
- [125] W. Buchmuller, V. Domcke, and K. Schmitz, *Metastable cosmic strings*, *JCAP* **11** (2023) 020, [[2307.04691](#)].
- [126] S. Blasi, V. Brdar, and K. Schmitz, *Has NANOGrav found first evidence for cosmic strings?*, *Phys. Rev. Lett.* **126** (2021), no. 4

- 041305, [[2009.06607](#)].
- [127] J. Ellis and M. Lewicki, *Cosmic String Interpretation of NANOGrav Pulsar Timing Data*, *Phys. Rev. Lett.* **126** (2021), no. 4 041304, [[2009.06555](#)].
 - [128] W. Buchmuller, V. Domcke, and K. Schmitz, *From NANOGrav to LIGO with metastable cosmic strings*, *Phys. Lett. B* **811** (2020) 135914, [[2009.10649](#)].
 - [129] Y. Gouttenoire and E. Vitagliano, *Domain wall interpretation of the PTA signal confronting black hole overproduction*, [[2306.17841](#)].
 - [130] S. Vagnozzi, *Inflationary interpretation of the stochastic gravitational wave background signal detected by pulsar timing array experiments*, *JHEAp* **39** (2023) 81–98, [[2306.16912](#)].
 - [131] S. Vagnozzi, *Implications of the NANOGrav results for inflation*, *Mon. Not. Roy. Astron. Soc.* **502** (2021), no. 1 L11–L15, [[2009.13432](#)].
 - [132] A. Ashoorioon, K. Rezazadeh, and A. Rostami, *NANOGrav signal from the end of inflation and the LIGO mass and heavier primordial black holes*, *Phys. Lett. B* **835** (2022) 137542, [[2202.01131](#)].
 - [133] V. Dandoy, V. Domcke, and F. Rompineve, *Search for scalar induced gravitational waves in the international pulsar timing array data release 2 and NANOgrav 12.5 years datasets*, *SciPost Phys. Core* **6** (2023) 060, [[2302.07901](#)].
 - [134] V. De Luca, G. Franciolini, and A. Riotto, *NANOGrav Data Hints at Primordial Black Holes as Dark Matter*, *Phys. Rev. Lett.* **126** (2021), no. 4 041303, [[2009.08268](#)].
 - [135] K. Kohri and T. Terada, *Solar-Mass Primordial Black Holes Explain NANOGrav Hint of Gravitational Waves*, *Phys. Lett. B* **813** (2021) 136040, [[2009.11853](#)].
 - [136] V. Vaskonen and H. Veermäe, *Did NANOGrav see a signal from primordial black hole formation?*, *Phys. Rev. Lett.* **126** (2021), no. 5 051303, [[2009.07832](#)].
 - [137] C. Ünal, E. D. Kovetz, and S. P. Patil, *Multimessenger probes of inflationary fluctuations and primordial black holes*, *Phys. Rev. D* **103** (2021), no. 6 063519, [[2008.11184](#)].
 - [138] A. Roper Pol, C. Caprini, A. Neronov, and D. Semikoz, *Gravitational wave signal from primordial magnetic fields in the Pulsar Timing Array frequency band*, *Phys. Rev. D* **105** (2022), no. 12 123502, [[2201.05630](#)].

- [139] A. Ghoshal and A. Strumia, *Traversing a kinetic pole during inflation: primordial black holes and gravitational waves*, [[2311.16236](#)].
- [140] J. Kormendy and D. Richstone, *Inward bound: The Search for supermassive black holes in galactic nuclei*, *Ann. Rev. Astron. Astrophys.* **33** (1995) 581.
- [141] J. Magorrian et al., *The Demography of massive dark objects in galaxy centers*, *Astron. J.* **115** (1998) 2285, [[astro-ph/9708072](#)].
- [142] **Event Horizon Telescope Collaboration**, K. Akiyama et al., *First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole*, *Astrophys. J. Lett.* **875** (2019) L1, [[1906.11238](#)].
- [143] A. M. Ghez, B. L. Klein, M. Morris, and E. E. Becklin, *High proper motion stars in the vicinity of Sgr A*: Evidence for a supermassive black hole at the center of our galaxy*, *Astrophys. J.* **509** (1998) 678–686, [[astro-ph/9807210](#)].
- [144] A. Eckart and R. Genzel, *Observations of stellar proper motions near the Galactic Centre*, *Nature* **383** (1996) 415–417.
- [145] M. Milosavljevic and D. Merritt, *The Final parsec problem*, *AIP Conf. Proc.* **686** (2003), no. 1 201–210, [[astro-ph/0212270](#)].
- [146] A. De Rosa et al., *The quest for dual and binary supermassive black holes: A multi-messenger view*, *New Astron. Rev.* **86** (2019) 101525, [[2001.06293](#)].
- [147] C. Rodriguez, G. B. Taylor, R. T. Zavala, A. B. Peck, L. K. Pollack, et al., *A Compact Supermassive Binary Black Hole System*, *Astrophys. J.* **646** (2006) 49–60, [[astro-ph/0604042](#)].
- [148] G. Alonso-Álvarez, J. M. Cline, and C. Dewar, *Self-interacting dark matter solves the final parsec problem of supermassive black hole mergers*, [[2401.14450](#)].
- [149] **NANOGrav Collaboration**, G. Agazie et al., *The NANOGrav 15 yr Data Set: Constraints on Supermassive Black Hole Binaries from the Gravitational-wave Background*, *Astrophys. J. Lett.* **952** (2023), no. 2 L37, [[2306.16220](#)].
- [150] E. S. Phinney, *A Practical theorem on gravitational wave backgrounds*, [[astro-ph/0108028](#)].
- [151] M. B. Hindmarsh, M. Lüben, J. Lumma, and M. Pauly, *Phase transitions in the early universe*, *SciPost Phys. Lect. Notes* **24** (2021) 1, [[2008.09136](#)].
- [152] R. Kubo, *Statistical mechanical theory of irreversible processes. 1. General theory and simple applications in magnetic and conduction problems*, *J. Phys. Soc. Jap.* **12** (1957) 570–586.

- [153] P. C. Martin and J. S. Schwinger, *Theory of many particle systems. 1.*, *Phys. Rev.* **115** (1959) 1342–1373.
- [154] T. Matsubara, *A New approach to quantum statistical mechanics*, *Prog. Theor. Phys.* **14** (1955) 351–378.
- [155] L. V. Keldysh, *Diagram technique for nonequilibrium processes*, *Zh. Eksp. Teor. Fiz.* **47** (1964) 1515–1527.
- [156] M. Quiros, *Finite temperature field theory and phase transitions*, in *ICTP Summer School in High-Energy Physics and Cosmology*, pp. 187–259, 1, 1999. [[hep-ph/9901312](#)].
- [157] P. Ilten, Y. Soreq, M. Williams, and W. Xue, *Serendipity in dark photon searches*, *JHEP* **06** (2018) 004, [[1801.04847](#)].
- [158] M. Bauer, P. Foldenauer, and J. Jaeckel, *Hunting All the Hidden Photons*, *JHEP* **07** (2018) 094, [[1803.05466](#)].
- [159] T. Ferber, A. Grohsjean, and F. Kahlhoefer, *Dark Higgs Bosons at Colliders*, [[2305.16169](#)].
- [160] S. Weinberg, *The quantum theory of fields. Vol. 2: Modern applications*. Cambridge University Press, 8, 2013.
- [161] S. R. Coleman and E. J. Weinberg, *Radiative Corrections as the Origin of Spontaneous Symmetry Breaking*, *Phys. Rev. D* **7** (1973) 1888–1910.
- [162] C. Delaunay, C. Grojean, and J. D. Wells, *Dynamics of Non-renormalizable Electroweak Symmetry Breaking*, *JHEP* **04** (2008) 029, [[0711.2511](#)].
- [163] D. Curtin, P. Meade, and H. Ramani, *Thermal Resummation and Phase Transitions*, *Eur. Phys. J. C* **78** (2018), no. 9 787, [[1612.00466](#)].
- [164] P. Basler and M. Mühlleitner, *BSMPT (Beyond the Standard Model Phase Transitions): A tool for the electroweak phase transition in extended Higgs sectors*, *Comput. Phys. Commun.* **237** (2019) 62–85, [[1803.02846](#)].
- [165] P. Athron, C. Balázs, A. Fowlie, L. Morris, and L. Wu, *Cosmological phase transitions: from perturbative particle physics to gravitational waves*, [[2305.02357](#)].
- [166] M. Laine and A. Vuorinen, *Basics of Thermal Field Theory*, vol. 925. Springer, 2016.
- [167] P. B. Arnold and O. Espinosa, *The Effective potential and first order phase transitions: Beyond leading-order*, *Phys. Rev. D* **47** (1993) 3546, [[hep-ph/9212235](#)]. [Erratum: Phys.Rev.D 50, 6662 (1994)].

- [168] R. R. Parwani, *Resummation in a hot scalar field theory*, *Phys. Rev. D* **45** (1992) 4695, [[hep-ph/9204216](#)]. [Erratum: Phys.Rev.D 48, 5965 (1993)].
- [169] M. E. Carrington, *The Effective potential at finite temperature in the Standard Model*, *Phys. Rev. D* **45** (1992) 2933–2944.
- [170] M. J. Baker, M. Breitbach, J. Kopp, and L. Mittnacht, *Dynamic Freeze-In: Impact of Thermal Masses and Cosmological Phase Transitions on Dark Matter Production*, *JHEP* **03** (2018) 114, [[1712.03962](#)].
- [171] A. Ekstedt, P. Schicho, and T. V. I. Tenkanen, *DRalgo: A package for effective field theory approach for thermal phase transitions*, *Comput. Phys. Commun.* **288** (2023) 108725, [[2205.08815](#)].
- [172] A. D. Linde, *Fate of the False Vacuum at Finite Temperature: Theory and Applications*, *Phys. Lett. B* **100** (1981) 37–40.
- [173] A. D. Linde, *Decay of the False Vacuum at Finite Temperature*, *Nucl. Phys. B* **216** (1983) 421. [Erratum: Nucl.Phys.B 223, 544 (1983)].
- [174] S. R. Coleman, *The Fate of the False Vacuum. 1. Semiclassical Theory*, *Phys. Rev. D* **15** (1977) 2929–2936. [Erratum: Phys.Rev.D 16, 1248 (1977)].
- [175] A. Ekstedt, O. Gould, and J. Hirvonen, *BubbleDet: a Python package to compute functional determinants for bubble nucleation*, *JHEP* **12** (2023) 056, [[2308.15652](#)].
- [176] H.-K. Guo, K. Sinha, D. Vagie, and G. White, *The benefits of diligence: how precise are predicted gravitational wave spectra in models with phase transitions?*, *JHEP* **06** (2021) 164, [[2103.06933](#)].
- [177] P. Athron, L. Morris, and Z. Xu, *How robust are gravitational wave predictions from cosmological phase transitions?*, [[2309.05474](#)].
- [178] C. Caprini et al., *Detecting gravitational waves from cosmological phase transitions with LISA: an update*, *JCAP* **03** (2020) 024, [[1910.13125](#)].
- [179] J. Ellis, M. Lewicki, and J. M. No, *On the Maximal Strength of a First-Order Electroweak Phase Transition and its Gravitational Wave Signal*, *JCAP* **04** (2019) 003, [[1809.08242](#)].
- [180] M. Hindmarsh, S. J. Huber, K. Rummukainen, and D. J. Weir, *Numerical simulations of acoustically generated gravitational waves at a first order phase transition*, *Phys. Rev. D* **92** (2015), no. 12 123009, [[1504.03291](#)].

- [181] P. Athron et al., *Global fits of axion-like particles to XENON1T and astrophysical data*, *JHEP* **05** (2021) 159, [[2007.05517](#)].
- [182] F. Giese, T. Konstandin, and J. van de Vis, *Model-independent energy budget of cosmological first-order phase transitions—A sound argument to go beyond the bag model*, *JCAP* **07** (2020), no. 07 057, [[2004.06995](#)].
- [183] L. Randall and G. Servant, *Gravitational waves from warped spacetime*, *JHEP* **05** (2007) 054, [[hep-ph/0607158](#)].
- [184] T. Konstandin and G. Servant, *Cosmological Consequences of Nearly Conformal Dynamics at the TeV scale*, *JCAP* **12** (2011) 009, [[1104.4791](#)].
- [185] J. Ellis, M. Lewicki, and V. Vaskonen, *Updated predictions for gravitational waves produced in a strongly supercooled phase transition*, *JCAP* **11** (2020) 020, [[2007.15586](#)].
- [186] M. Kierkla, A. Karam, and B. Swiezewska, *Conformal model for gravitational waves and dark matter: a status update*, *JHEP* **03** (2023) 007, [[2210.07075](#)].
- [187] B. von Harling and G. Servant, *QCD-induced Electroweak Phase Transition*, *JHEP* **01** (2018) 159, [[1711.11554](#)].
- [188] S. Bruggisser, B. Von Harling, O. Matsedonskyi, and G. Servant, *Electroweak Phase Transition and Baryogenesis in Composite Higgs Models*, *JHEP* **12** (2018) 099, [[1804.07314](#)].
- [189] S. Bruggisser, B. von Harling, O. Matsedonskyi, and G. Servant, *Status of electroweak baryogenesis in minimal composite Higgs*, *JHEP* **08** (2023) 012, [[2212.11953](#)].
- [190] I. Baldes, Y. Gouttenoire, F. Sala, and G. Servant, *Supercool composite Dark Matter beyond 100 TeV*, *JHEP* **07** (2022) 084, [[2110.13926](#)].
- [191] A. Roper Pol, S. Mandal, A. Brandenburg, T. Kahnashvili, and A. Kosowsky, *Numerical simulations of gravitational waves from early-universe turbulence*, *Phys. Rev. D* **102** (2020), no. 8 083512, [[1903.08585](#)].
- [192] J. R. Espinosa, T. Konstandin, J. M. No, and G. Servant, *Energy Budget of Cosmological First-order Phase Transitions*, *JCAP* **06** (2010) 028, [[1004.4187](#)].
- [193] M. Kamionkowski, A. Kosowsky, and M. S. Turner, *Gravitational radiation from first order phase transitions*, *Phys. Rev. D* **49** (1994) 2837–2851, [[astro-ph/9310044](#)].
- [194] A. Kosowsky and M. S. Turner, *Gravitational radiation from colliding vacuum bubbles: envelope approximation to many bubble collisions*, *Phys. Rev. D* **47** (1993) 4372–4391,

- [astro-ph/9211004].
- [195] A. Kosowsky, M. S. Turner, and R. Watkins, *Gravitational radiation from colliding vacuum bubbles*, *Phys. Rev. D* **45** (1992) 4514–4535.
 - [196] R. Jinno and M. Takimoto, *Gravitational waves from bubble collisions: An analytic derivation*, *Phys. Rev. D* **95** (2017), no. 2 024009, [[1605.01403](#)].
 - [197] C. Caprini, R. Durrer, and G. Servant, *Gravitational wave generation from bubble collisions in first-order phase transitions: An analytic approach*, *Phys. Rev. D* **77** (2008) 124015, [[0711.2593](#)].
 - [198] S. J. Huber and T. Konstandin, *Gravitational Wave Production by Collisions: More Bubbles*, *JCAP* **09** (2008) 022, [[0806.1828](#)].
 - [199] D. J. Weir, *Revisiting the envelope approximation: gravitational waves from bubble collisions*, *Phys. Rev. D* **93** (2016), no. 12 124037, [[1604.08429](#)].
 - [200] T. Konstandin, *Gravitational radiation from a bulk flow model*, *JCAP* **03** (2018) 047, [[1712.06869](#)].
 - [201] R. Durrer and C. Caprini, *Primordial magnetic fields and causality*, *JCAP* **11** (2003) 010, [[astro-ph/0305059](#)].
 - [202] C. Caprini, R. Durrer, T. Konstandin, and G. Servant, *General Properties of the Gravitational Wave Spectrum from Phase Transitions*, *Phys. Rev. D* **79** (2009) 083519, [[0901.1661](#)].
 - [203] R.-G. Cai, S. Pi, and M. Sasaki, *Universal infrared scaling of gravitational wave background spectra*, *Phys. Rev. D* **102** (2020), no. 8 083528, [[1909.13728](#)].
 - [204] D. Cutting, M. Hindmarsh, and D. J. Weir, *Gravitational waves from vacuum first-order phase transitions: from the envelope to the lattice*, *Phys. Rev. D* **97** (2018), no. 12 123513, [[1802.05712](#)].
 - [205] D. Cutting, E. G. Escartin, M. Hindmarsh, and D. J. Weir, *Gravitational waves from vacuum first order phase transitions II: from thin to thick walls*, *Phys. Rev. D* **103** (2021), no. 2 023531, [[2005.13537](#)].
 - [206] R. Jinno and M. Takimoto, *Gravitational waves from bubble dynamics: Beyond the Envelope*, *JCAP* **01** (2019) 060, [[1707.03111](#)].
 - [207] Y. Gouttenoire, *Beyond the Standard Model Cocktail*. Springer Theses. Springer, Cham, 2022.
 - [208] M. E. A. Vanvlasselaer, *Dynamics of phase transitions in the early universe and cosmological consequences*. PhD thesis, SISSA,

- Trieste, 2022.
- [209] M. Hindmarsh, S. J. Huber, K. Rummukainen, and D. J. Weir, *Gravitational waves from the sound of a first order phase transition*, *Phys. Rev. Lett.* **112** (2014) 041301, [[1304.2433](#)].
 - [210] J. T. Giblin and J. B. Mertens, *Gravitational radiation from first-order phase transitions in the presence of a fluid*, *Phys. Rev. D* **90** (2014), no. 2 023532, [[1405.4005](#)].
 - [211] D. Cutting, M. Hindmarsh, and D. J. Weir, *Vorticity, kinetic energy, and suppressed gravitational wave production in strong first order phase transitions*, *Phys. Rev. Lett.* **125** (2020), no. 2 021302, [[1906.00480](#)].
 - [212] M. Hindmarsh, S. J. Huber, K. Rummukainen, and D. J. Weir, *Shape of the acoustic gravitational wave power spectrum from a first order phase transition*, *Phys. Rev. D* **96** (2017), no. 10 103520, [[1704.05871](#)]. [Erratum: Phys.Rev.D 101, 089902 (2020)].
 - [213] M. Hindmarsh, *Sound shell model for acoustic gravitational wave production at a first-order phase transition in the early Universe*, *Phys. Rev. Lett.* **120** (2018), no. 7 071301, [[1608.04735](#)].
 - [214] M. Hindmarsh and M. Hijazi, *Gravitational waves from first order cosmological phase transitions in the Sound Shell Model*, *JCAP* **12** (2019) 062, [[1909.10040](#)].
 - [215] H.-K. Guo, K. Sinha, D. Vagie, and G. White, *Phase Transitions in an Expanding Universe: Stochastic Gravitational Waves in Standard and Non-Standard Histories*, *JCAP* **01** (2021) 001, [[2007.08537](#)].
 - [216] J. A. Casey-Clyde, C. M. F. Mingarelli, J. E. Greene, K. Pardo, M. Nañez, et al., *A Quasar-based Supermassive Black Hole Binary Population Model: Implications for the Gravitational Wave Background*, *Astrophys. J.* **924** (2022), no. 2 93, [[2107.11390](#)].
 - [217] L. Z. Kelley, L. Blecha, and L. Hernquist, *Massive Black Hole Binary Mergers in Dynamical Galactic Environments*, *Mon. Not. Roy. Astron. Soc.* **464** (2017), no. 3 3131–3157, [[1606.01900](#)].
 - [218] L. Z. Kelley, L. Blecha, L. Hernquist, A. Sesana, and S. R. Taylor, *The Gravitational Wave Background from Massive Black Hole Binaries in Illustris: spectral features and time to detection with pulsar timing arrays*, *Mon. Not. Roy. Astron. Soc.* **471** (2017), no. 4 4508–4526, [[1702.02180](#)].
 - [219] H. Middleton, A. Sesana, S. Chen, A. Vecchio, W. Del Pozzo, et al., *Retracted: Correction to: Massive black hole binary systems and the NANOGrav 12.5 yr results*, *Mon. Not. Roy. Astron. Soc.* **502** (2021), no. 1 L99–L103, [[2011.01246](#)]. [Erratum: Mon.Not.Roy.Astron.Soc. 526, L34 (2023)].

- [220] D. Izquierdo-Villalba, A. Sesana, S. Bonoli, and M. Colpi, *Massive black hole evolution models confronting the n-Hz amplitude of the stochastic gravitational wave background*, *Mon. Not. Roy. Astron. Soc.* **509** (2021), no. 3 3488–3503, [[2108.11671](#)].
- [221] M. Curyło and T. Bulik, *Predictions for LISA and PTA based on SHARK galaxy simulations*, *Astron. Astrophys.* **660** (2022) A68, [[2108.11232](#)].
- [222] J. J. Somalwar and V. Ravi, *The origin of the nano-Hertz stochastic gravitational wave background: the contribution from $z \gtrsim 1$ supermassive black-hole binaries*, [[2306.00898](#)].
- [223] S. Gori et al., *Dark Sector Physics at High-Intensity Experiments*, [[2209.04671](#)].
- [224] M. Pospelov, A. Ritz, and M. B. Voloshin, *Secluded WIMP Dark Matter*, *Phys. Lett. B* **662** (2008) 53–61, [[0711.4866](#)].
- [225] J. L. Feng, H. Tu, and H.-B. Yu, *Thermal Relics in Hidden Sectors*, *JCAP* **10** (2008) 043, [[0808.2318](#)].
- [226] M. Pospelov, *Secluded U(1) below the weak scale*, *Phys. Rev. D* **80** (2009) 095002, [[0811.1030](#)].
- [227] M. Hufnagel, K. Schmidt-Hoberg, and S. Wild, *BBN constraints on MeV-scale dark sectors. Part II. Electromagnetic decays*, *JCAP* **11** (2018) 032, [[1808.09324](#)].
- [228] L. Forestell, D. E. Morrissey, and G. White, *Limits from BBN on Light Electromagnetic Decays*, *JHEP* **01** (2019) 074, [[1809.01179](#)].
- [229] P. F. Depta, *Dark Sector Cosmologies: Evolution and Constraints*. PhD thesis, Hamburg U., Hamburg, 2021.
- [230] P. F. Depta, M. Hufnagel, and K. Schmidt-Hoberg, *ACROPOLIS: A generiC fRamework fOr Photodisintegration Of LIght elements*, *JCAP* **03** (2021) 061, [[2011.06518](#)].
- [231] M. Kawasaki, K. Kohri, T. Moroi, K. Murai, and H. Murayama, *Big-bang nucleosynthesis with sub-GeV massive decaying particles*, *JCAP* **12** (2020) 048, [[2006.14803](#)].
- [232] Y. Bai and M. Korwar, *Cosmological constraints on first-order phase transitions*, *Phys. Rev. D* **105** (2022), no. 9 095015, [[2109.14765](#)].
- [233] S. Deng and L. Bian, *Constraining low-scale dark phase transitions with cosmological observations*, [[2304.06576](#)].
- [234] T. V. I. Tenkanen and J. van de Vis, *Speed of sound in cosmological phase transitions and effect on gravitational waves*, *JHEP* **08** (2022) 302, [[2206.01130](#)].

- [235] K. Freese and M. W. Winkler, *Have pulsar timing arrays detected the hot big bang: Gravitational waves from strong first order phase transitions in the early Universe*, *Phys. Rev. D* **106** (2022), no. 10 103523, [[2208.03330](#)].
- [236] R. Jinno, T. Konstandin, H. Rubira, and I. Stomberg, *Higgsless simulations of cosmological phase transitions and gravitational waves*, *JCAP* **02** (2023) 011, [[2209.04369](#)].
- [237] R. Jinno, T. Konstandin, and H. Rubira, *A hybrid simulation of gravitational wave production in first-order phase transitions*, *JCAP* **04** (2021) 014, [[2010.00971](#)].
- [238] **NANOGrav Collaboration**, Z. Arzoumanian et al., *The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background*, *Astrophys. J. Lett.* **905** (2020), no. 2 L34, [[2009.04496](#)].
- [239] M. Benetti, L. L. Graef, and S. Vagnozzi, *Primordial gravitational waves from NANOGrav: A broken power-law approach*, *Phys. Rev. D* **105** (2022), no. 4 043520, [[2111.04758](#)].
- [240] D. Wang, *Novel Physics with International Pulsar Timing Array: Axionlike Particles, Domain Walls and Cosmic Strings*, [[2203.10959](#)].
- [241] D. Wang, *Squeezing Cosmological Phase Transitions with International Pulsar Timing Array*, [[2201.09295](#)].
- [242] W. Ratzinger, *Misaligned, tilted and distorted: the hard life of audible axions*. PhD thesis, Mainz U., 2023.
- [243] S. Taylor, S. Vigeland, J. Simon, B. Becksy, and A. Johnson, *Pta gwb analysis*, 2021.
- [244] P. Virtanen et al., *SciPy 1.0—Fundamental Algorithms for Scientific Computing in Python*, *Nature Meth.* **17** (2020) 261, [[1907.10121](#)].
- [245] S. R. Hinton, *ChainConsumer*, *The Journal of Open Source Software* **1** (Aug., 2016) 00045.
- [246] S. Hee, W. Handley, M. P. Hobson, and A. N. Lasenby, *Bayesian model selection without evidences: application to the dark energy equation-of-state*, *Mon. Not. Roy. Astron. Soc.* **455** (2016), no. 3 2461–2473, [[1506.09024](#)].
- [247] S. J. Godsill, *On the relationship between markov chain monte carlo methods for model uncertainty*, *Journal of Computational and Graphical Statistics* **10** (2001), no. 2 230–248.
- [248] S. J. Chamberlin, J. D. E. Creighton, X. Siemens, P. Demorest, J. Ellis, et al., *Time-domain Implementation of the Optimal Cross-Correlation Statistic for Stochastic Gravitational-Wave*

- Background Searches in Pulsar Timing Data,* *Phys. Rev. D* **91** (2015), no. 4 044048, [[1410.8256](#)].
- [249] B. P. Carlin and S. Chib, *Bayesian model choice via markov chain monte carlo methods,* *Journal of the Royal Statistical Society. Series B (Methodological)* **57** (1995), no. 3 473–484.
- [250] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, et al., *The Cosmic Microwave Background spectrum from the full COBE FIRAS data set,* *Astrophys. J.* **473** (1996) 576, [[astro-ph/9605054](#)].
- [251] P. F. Depta, M. Hufnagel, and K. Schmidt-Hoberg, *Updated BBN constraints on electromagnetic decays of MeV-scale particles,* *JCAP* **04** (2021) 011, [[2011.06519](#)].
- [252] N. Ramberg, W. Ratzinger, and P. Schwaller, *One μ to rule them all: CMB spectral distortions can probe domain walls, cosmic strings and low scale phase transitions,* *JCAP* **02** (2023) 039, [[2209.14313](#)].
- [253] J. Liu, L. Bian, R.-G. Cai, Z.-K. Guo, and S.-J. Wang, *Constraining First-Order Phase Transitions with Curvature Perturbations,* *Phys. Rev. Lett.* **130** (2023), no. 5 051001, [[2208.14086](#)].
- [254] M. Lewicki, P. Toczek, and V. Vaskonen, *Primordial black holes from strong first-order phase transitions,* *JHEP* **09** (2023) 092, [[2305.04924](#)].
- [255] M. J. Baker, M. Breitbach, J. Kopp, and L. Mittnacht, *Primordial Black Holes from First-Order Cosmological Phase Transitions,* [[2105.07481](#)].
- [256] M. W. Winkler, *Decay and detection of a light scalar boson mixing with the Higgs boson,* *Phys. Rev. D* **99** (2019), no. 1 015018, [[1809.01876](#)].
- [257] J. Antoniadis et al., *The International Pulsar Timing Array second data release: Search for an isotropic gravitational wave background,* *Mon. Not. Roy. Astron. Soc.* **510** (2022), no. 4 4873–4887, [[2201.03980](#)].
- [258] F. Giese, T. Konstandin, K. Schmitz, and J. van de Vis, *Model-independent energy budget for LISA,* *JCAP* **01** (2021) 072, [[2010.09744](#)].
- [259] M. W. Winkler and K. Freese, *Origin of the Stochastic Gravitational Wave Background: First-Order Phase Transition vs. Black Hole Mergers,* [[2401.13729](#)].
- [260] F. Schulze, L. Valbusa Dall’Armi, J. Lesgourgues, A. Ricciardone, N. Bartolo, et al., *GW_CLASS: Cosmological Gravitational Wave Background in the cosmic linear anisotropy solving system,*

- JCAP* **10** (2023) 025, [[2305.01602](#)].
- [261] **LISA Cosmology Working Group Collaboration**, N. Bartolo et al., *Probing anisotropies of the Stochastic Gravitational Wave Background with LISA*, *JCAP* **11** (2022) 009, [[2201.08782](#)].
 - [262] S. R. Taylor, R. van Haasteren, and A. Sesana, *From Bright Binaries To Bumpy Backgrounds: Mapping Realistic Gravitational Wave Skies With Pulsar-Timing Arrays*, *Phys. Rev. D* **102** (2020), no. 8 084039, [[2006.04810](#)].
 - [263] **NANOGrav Collaboration**, Z. Arzoumanian et al., *The NANOGrav 12.5-year Data Set: Bayesian Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries*, [[2301.03608](#)].
 - [264] **IPTA Collaboration**, M. Falxa et al., *Searching for continuous Gravitational Waves in the second data release of the International Pulsar Timing Array*, *Mon. Not. Roy. Astron. Soc.* **521** (2023), no. 4 5077–5086, [[2303.10767](#)].
 - [265] G. Agazie et al., *The NANOGrav 15 yr Data Set: Looking for Signs of Discreteness in the Gravitational-wave Background*, [[2404.07020](#)].
 - [266] R. Kato and J. Soda, *Probing circular polarization in stochastic gravitational wave background with pulsar timing arrays*, *Phys. Rev. D* **93** (2016), no. 6 062003, [[1512.09139](#)].
 - [267] C. Conneely, A. H. Jaffe, and C. M. F. Mingarelli, *On the Amplitude and Stokes Parameters of a Stochastic Gravitational-Wave Background*, *Mon. Not. Roy. Astron. Soc.* **487** (2019), no. 1 562–579, [[1808.05920](#)].
 - [268] S. C. Hotinli, M. Kamionkowski, and A. H. Jaffe, *The search for anisotropy in the gravitational-wave background with pulsar-timing arrays*, *Open J. Astrophys.* **2** (2019), no. 1 8, [[1904.05348](#)].
 - [269] E. Belgacem and M. Kamionkowski, *Chirality of the gravitational-wave background and pulsar-timing arrays*, *Phys. Rev. D* **102** (2020), no. 2 023004, [[2004.05480](#)].
 - [270] G. Sato-Polito and M. Kamionkowski, *Pulsar-timing measurement of the circular polarization of the stochastic gravitational-wave background*, *Phys. Rev. D* **106** (2022), no. 2 023004, [[2111.05867](#)].
 - [271] L. Valbusa Dall’Armi, A. Nishizawa, A. Ricciardone, and S. Matarrese, *Circular Polarization of the Astrophysical Gravitational Wave Background*, [[2301.08205](#)].
 - [272] J. Ellis, M. Fairbairn, G. Hütsi, M. Raidal, J. Urrutia, et al., *Prospects for Future Binary Black Hole GW Studies in Light of*

- PTA Measurements*, [[2301.13854](#)].
- [273] **NANOGrav Collaboration**, N. S. Pol et al., *Astrophysics Milestones for Pulsar Timing Array Gravitational-wave Detection*, *Astrophys. J. Lett.* **911** (2021), no. 2 L34, [[2010.11950](#)].
- [274] K. Abazajian et al., *CMB-S4 Science Case, Reference Design, and Project Plan*, [[1907.04473](#)].
- [275] **Simons Observatory Collaboration**, P. Ade et al., *The Simons Observatory: Science goals and forecasts*, *JCAP* **02** (2019) 056, [[1808.07445](#)].
- [276] C. Dvorkin et al., *The Physics of Light Relics*, in *Snowmass 2021*, 3, 2022. [[2203.07943](#)].
- [277] J. Chluba et al., *New horizons in cosmology with spectral distortions of the cosmic microwave background*, *Exper. Astron.* **51** (2021), no. 3 1515–1554, [[1909.01593](#)].
- [278] B. Batell, M. Pospelov, and A. Ritz, *Exploring Portals to a Hidden Sector Through Fixed Targets*, *Phys. Rev. D* **80** (2009) 095024, [[0906.5614](#)].
- [279] T. Bringmann, P. F. Depta, M. Hufnagel, and K. Schmidt-Hoberg, *Precise dark matter relic abundance in decoupled sectors*, *Phys. Lett. B* **817** (2021) 136341, [[2007.03696](#)].
- [280] M. Hufnagel and M. H. G. Tytgat, *The domain of a cannibal dark matter*, *JCAP* **09** (2023) 012, [[2212.09759](#)].
- [281] A. Fowlie, *Bayesian and frequentist approaches to resonance searches*, *JINST* **14** (2019), no. 10 P10031, [[1902.03243](#)].
- [282] **LISA Collaboration**, P. Amaro-Seoane et al., *Laser Interferometer Space Antenna*, [[1702.00786](#)].
- [283] J. R. Espinosa, T. Konstandin, J. M. No, and M. Quiros, *Some Cosmological Implications of Hidden Sectors*, *Phys. Rev. D* **78** (2008) 123528, [[0809.3215](#)].
- [284] T. Alanen, K. Tuominen, and V. Vaskonen, *Strong phase transition, dark matter and vacuum stability from simple hidden sectors*, *Nucl. Phys. B* **889** (2014) 692–711, [[1407.0688](#)].
- [285] I. Baldes, *Gravitational waves from the asymmetric-dark-matter generating phase transition*, *JCAP* **05** (2017) 028, [[1702.02117](#)].
- [286] E. Hall, T. Konstandin, R. McGehee, and H. Murayama, *Asymmetric matter from a dark first-order phase transition*, *Phys. Rev. D* **107** (2023), no. 5 055011, [[1911.12342](#)].
- [287] M. J. Baker, J. Kopp, and A. J. Long, *Filtered Dark Matter at a First Order Phase Transition*, *Phys. Rev. Lett.* **125** (2020), no. 15 151102, [[1912.02830](#)].

- [288] A. Azatov, M. Vanvlasselaer, and W. Yin, *Dark Matter production from relativistic bubble walls*, *JHEP* **03** (2021) 288, [[2101.05721](#)].
- [289] M. Lewicki, K. Müürsepp, J. Pata, M. Vasar, V. Vaskonen, et al., *Dynamics of false vacuum bubbles with trapped particles*, *Phys. Rev. D* **108** (2023), no. 3 036023, [[2305.07702](#)].
- [290] S. Kanemura and S.-P. Li, *Dark phase transition from WIMP: complementary tests from gravitational waves and colliders*, [[2308.16390](#)].
- [291] S. Biondini, P. Schicho, and T. V. I. Tenkanen, *Strong electroweak phase transition in t-channel simplified dark matter models*, *JCAP* **10** (2022) 044, [[2207.12207](#)].
- [292] G. Arcadi, G. C. Dorsch, J. P. Neto, F. S. Queiroz, and Y. M. Oviedo-Torres, *Probing a dark sector with collider physics, direct detection, and gravitational waves*, *Phys. Lett. B* **848** (2024) 138382, [[2307.06376](#)].
- [293] E. Madge and P. Schwaller, *Leptophilic dark matter from gauged lepton number: Phenomenology and gravitational wave signatures*, *JHEP* **02** (2019) 048, [[1809.09110](#)].
- [294] D. Croon, T. E. Gonzalo, and G. White, *Gravitational Waves from a Pati-Salam Phase Transition*, *JHEP* **02** (2019) 083, [[1812.02747](#)].
- [295] P. Athron, C. Balázs, T. E. Gonzalo, and M. Pearce, *Falsifying Pati-Salam models with LIGO*, [[2307.02544](#)].
- [296] S. Weinberg, *Mass of the Higgs Boson*, *Phys. Rev. Lett.* **36** (1976) 294–296.
- [297] M. Fairbairn, E. Hardy, and A. Wickens, *Hearing without seeing: gravitational waves from hot and cold hidden sectors*, *JHEP* **07** (2019) 044, [[1901.11038](#)].
- [298] S.-P. Li and K.-P. Xie, *Collider test of nano-Hertz gravitational waves from pulsar timing arrays*, *Phys. Rev. D* **108** (2023), no. 5 055018, [[2307.01086](#)].
- [299] C. L. Wainwright, *CosmoTransitions: Computing Cosmological Phase Transition Temperatures and Bubble Profiles with Multiple Fields*, *Comput. Phys. Commun.* **183** (2012) 2006–2013, [[1109.4189](#)].
- [300] D. Croon, O. Gould, P. Schicho, T. V. I. Tenkanen, and G. White, *Theoretical uncertainties for cosmological first-order phase transitions*, *JHEP* **04** (2021) 055, [[2009.10080](#)].
- [301] P. H. Ginsparg, *First Order and Second Order Phase Transitions in Gauge Theories at Finite Temperature*, *Nucl. Phys. B* **170**

- (1980) 388–408.
- [302] D. Bodeker and G. D. Moore, *Electroweak Bubble Wall Speed Limit*, *JCAP* **05** (2017) 025, [[1703.08215](#)].
- [303] Y. Gouttenoire, R. Jinno, and F. Sala, *Friction pressure on relativistic bubble walls*, *JHEP* **05** (2022) 004, [[2112.07686](#)].
- [304] A. Azatov, G. Barni, R. Petrossian-Byrne, and M. Vanvlasselaer, *Quantisation Across Bubble Walls and Friction*, [[2310.06972](#)].
- [305] S. S. AbdusSalam et al., *Simple and statistically sound recommendations for analysing physical theories*, *Rept. Prog. Phys.* **85** (2022), no. 5 052201, [[2012.09874](#)].
- [306] B. W. Lee and S. Weinberg, *Cosmological Lower Bound on Heavy Neutrino Masses*, *Phys. Rev. Lett.* **39** (1977) 165–168.
- [307] T. Bringmann, J. Edsjö, P. Gondolo, P. Ullio, and L. Bergström, *DarkSUSY 6 : An Advanced Tool to Compute Dark Matter Properties Numerically*, *JCAP* **07** (2018) 033, [[1802.03399](#)].
- [308] P. Gondolo and G. Gelmini, *Cosmic abundances of stable particles: Improved analysis*, *Nucl. Phys. B* **360** (1991) 145–179.
- [309] D. Chway, T. H. Jung, and C. S. Shin, *Dark matter filtering-out effect during a first-order phase transition*, *Phys. Rev. D* **101** (2020), no. 9 095019, [[1912.04238](#)].
- [310] D. Pappadopulo, J. T. Ruderman, and G. Trevisan, *Dark matter freeze-out in a nonrelativistic sector*, *Phys. Rev. D* **94** (2016), no. 3 035005, [[1602.04219](#)].
- [311] M. Farina, D. Pappadopulo, J. T. Ruderman, and G. Trevisan, *Phases of Cannibal Dark Matter*, *JHEP* **12** (2016) 039, [[1607.03108](#)].
- [312] R. M. Schabinger and J. D. Wells, *A Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the large hadron collider*, *Phys. Rev. D* **72** (2005) 093007, [[hep-ph/0509209](#)].
- [313] B. Patt and F. Wilczek, *Higgs-field portal into hidden sectors*, [[hep-ph/0605188](#)].
- [314] E. Weihs and J. Zurita, *Dark Higgs Models at the 7 TeV LHC*, *JHEP* **02** (2012) 041, [[1110.5909](#)].
- [315] M. Duerr, F. Kahlhoefer, K. Schmidt-Hoberg, T. Schwetz, and S. Vogl, *How to save the WIMP: global analysis of a dark matter model with two s-channel mediators*, *JHEP* **09** (2016) 042, [[1606.07609](#)].
- [316] T. Bringmann, S. Heeba, F. Kahlhoefer, and K. Vangnnes, *Freezing-in a hot bath: resonances, medium effects and phase*

- transitions,* *JHEP* **02** (2022) 110, [[2111.14871](#)].
- [317] E. Bernreuther, F. Kahlhoefer, M. Lucente, and A. Morandini, *Searching for dark radiation at the LHC*, *JHEP* **08** (2022) 169, [[2204.01759](#)]. [Erratum: JHEP 06, 071 (2023)].
- [318] J. A. Evans, S. Gori, and J. Shelton, *Looking for the WIMP Next Door*, *JHEP* **02** (2018) 100, [[1712.03974](#)].
- [319] A. Fradette, M. Pospelov, J. Pradler, and A. Ritz, *Cosmological beam dump: constraints on dark scalars mixed with the Higgs boson*, *Phys. Rev. D* **99** (2019), no. 7 075004, [[1812.07585](#)].
- [320] A. Berlin, D. Hooper, and G. Krnjaic, *Thermal Dark Matter From A Highly Decoupled Sector*, *Phys. Rev. D* **94** (2016), no. 9 095019, [[1609.02555](#)].
- [321] N. F. Bell, Y. Cai, and R. K. Leane, *Dark Forces in the Sky: Signals from Z' and the Dark Higgs*, *JCAP* **08** (2016) 001, [[1605.09382](#)].
- [322] **CTA Collaboration**, A. Acharyya et al., *Sensitivity of the Cherenkov Telescope Array to a dark matter signal from the Galactic centre*, *JCAP* **01** (2021) 057, [[2007.16129](#)].
- [323] B.-H. Liu, L. D. McLerran, and N. Turok, *Bubble nucleation and growth at a baryon number producing electroweak phase transition*, *Phys. Rev. D* **46** (1992) 2668–2688.
- [324] D. Bodeker and G. D. Moore, *Can electroweak bubble walls run away?*, *JCAP* **05** (2009) 009, [[0903.4099](#)].
- [325] A. Azatov and M. Vanvlasselaer, *Bubble wall velocity: heavy physics effects*, *JCAP* **01** (2021) 058, [[2010.02590](#)].
- [326] I. Garcia Garcia, G. Koszegi, and R. Petrossian-Byrne, *Reflections on bubble walls*, *JHEP* **09** (2023) 013, [[2212.10572](#)].
- [327] S. Höche, J. Kozaczuk, A. J. Long, J. Turner, and Y. Wang, *Towards an all-orders calculation of the electroweak bubble wall velocity*, *JCAP* **03** (2021) 009, [[2007.10343](#)].
- [328] M. Lewicki and V. Vaskonen, *Gravitational waves from bubble collisions and fluid motion in strongly supercooled phase transitions*, *Eur. Phys. J. C* **83** (2023), no. 2 109, [[2208.11697](#)].
- [329] J. Ellis, M. Lewicki, J. M. No, and V. Vaskonen, *Gravitational wave energy budget in strongly supercooled phase transitions*, *JCAP* **06** (2019) 024, [[1903.09642](#)].
- [330] G. Arcadi, O. Lebedev, S. Pokorski, and T. Toma, *Real Scalar Dark Matter: Relativistic Treatment*, *JHEP* **08** (2019) 050, [[1906.07659](#)].

- [331] V. De Romeri, D. Karamitros, O. Lebedev, and T. Toma, *Neutrino dark matter and the Higgs portal: improved freeze-in analysis*, *JHEP* **10** (2020) 137, [[2003.12606](#)].
- [332] D. Hooper, A. Ireland, G. Krnjaic, and A. Stebbins, *Supermassive Primordial Black Holes From Inflation*, [[2308.00756](#)].
- [333] V. Atal, A. Sanglas, and N. Triantafyllou, *NANOGrav signal as mergers of Stupendously Large Primordial Black Holes*, *JCAP* **06** (2021) 022, [[2012.14721](#)].
- [334] T. Bringmann, P. F. Depta, V. Domcke, and K. Schmidt-Hoberg, *Towards closing the window of primordial black holes as dark matter: The case of large clustering*, *Phys. Rev. D* **99** (2019), no. 6 063532, [[1808.05910](#)].
- [335] P. Ajith et al., *A Template bank for gravitational waveforms from coalescing binary black holes. I. Non-spinning binaries*, *Phys. Rev. D* **77** (2008) 104017, [[0710.2335](#)]. [Erratum: Phys.Rev.D 79, 129901 (2009)].
- [336] M. Raidal, V. Vaskonen, and H. Veermäe, *Gravitational Waves from Primordial Black Hole Mergers*, *JCAP* **09** (2017) 037, [[1707.01480](#)].
- [337] T. Nakamura, M. Sasaki, T. Tanaka, and K. S. Thorne, *Gravitational waves from coalescing black hole MACHO binaries*, *Astrophys. J. Lett.* **487** (1997) L139–L142, [[astro-ph/9708060](#)].
- [338] K. Ioka, T. Chiba, T. Tanaka, and T. Nakamura, *Black hole binary formation in the expanding universe: Three body problem approximation*, *Phys. Rev. D* **58** (1998) 063003, [[astro-ph/9807018](#)].
- [339] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, *Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914*, *Phys. Rev. Lett.* **117** (2016), no. 6 061101, [[1603.08338](#)]. [Erratum: Phys.Rev.Lett. 121, 059901 (2018)].
- [340] P. C. Peters, *Gravitational Radiation and the Motion of Two Point Masses*, *Phys. Rev.* **136** (1964) B1224–B1232.
- [341] Y. Ali-Haïmoud, E. D. Kovetz, and M. Kamionkowski, *Merger rate of primordial black-hole binaries*, *Phys. Rev. D* **96** (2017), no. 12 123523, [[1709.06576](#)].
- [342] M. Raidal, C. Spethmann, V. Vaskonen, and H. Veermäe, *Formation and Evolution of Primordial Black Hole Binaries in the Early Universe*, *JCAP* **02** (2019) 018, [[1812.01930](#)].
- [343] J. Antoniadis et al., *The second data release from the European Pulsar Timing Array IV. Search for continuous gravitational wave signals*, [[2306.16226](#)].

- [344] A. Sesana, A. Vecchio, and C. N. Colacino, *The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays*, *Mon. Not. Roy. Astron. Soc.* **390** (2008) 192, [[0804.4476](#)].
- [345] B. J. Carr and S. W. Hawking, *Black holes in the early Universe*, *Mon. Not. Roy. Astron. Soc.* **168** (1974) 399–415.
- [346] K. M. Belotsky, V. I. Dokuchaev, Y. N. Eroshenko, E. A. Esipova, M. Y. Khlopov, et al., *Clusters of primordial black holes*, *Eur. Phys. J. C* **79** (2019), no. 3 246, [[1807.06590](#)].
- [347] E. Cotner and A. Kusenko, *Primordial black holes from scalar field evolution in the early universe*, *Phys. Rev. D* **96** (2017), no. 10 103002, [[1706.09003](#)].
- [348] F. Ferrer, E. Masso, G. Panico, O. Pujolas, and F. Rompineve, *Primordial Black Holes from the QCD axion*, *Phys. Rev. Lett.* **122** (2019), no. 10 101301, [[1807.01707](#)].
- [349] Y. Ali-Haïmoud, *Correlation Function of High-Threshold Regions and Application to the Initial Small-Scale Clustering of Primordial Black Holes*, *Phys. Rev. Lett.* **121** (2018), no. 8 081304, [[1805.05912](#)].
- [350] V. Desjacques and A. Riotto, *Spatial clustering of primordial black holes*, *Phys. Rev. D* **98** (2018), no. 12 123533, [[1806.10414](#)].
- [351] S. Young and C. T. Byrnes, *Long-short wavelength mode coupling tightens primordial black hole constraints*, *Phys. Rev. D* **91** (2015), no. 8 083521, [[1411.4620](#)].
- [352] T. Matsubara, T. Terada, K. Kohri, and S. Yokoyama, *Clustering of primordial black holes formed in a matter-dominated epoch*, *Phys. Rev. D* **100** (2019), no. 12 123544, [[1909.04053](#)].
- [353] B. J. Carr and M. Sakellariadou, *Dynamical constraints on dark compact objects*, *Astrophys. J.* **516** (1999) 195–220.
- [354] Y. Inoue and A. Kusenko, *New X-ray bound on density of primordial black holes*, *JCAP* **10** (2017) 034, [[1705.00791](#)].
- [355] B. J. Carr and J. E. Lidsey, *Primordial black holes and generalized constraints on chaotic inflation*, *Phys. Rev. D* **48** (1993) 543–553.
- [356] A. Mitridate, D. Wright, R. von Eckardstein, T. Schröder, J. Nay, et al., *PTArcade*, [[2306.16377](#)].

EIDESSTATTLICHE VERSICHERUNG

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Hamburg, den 17. Oktober 2024

Carlo Tasillo