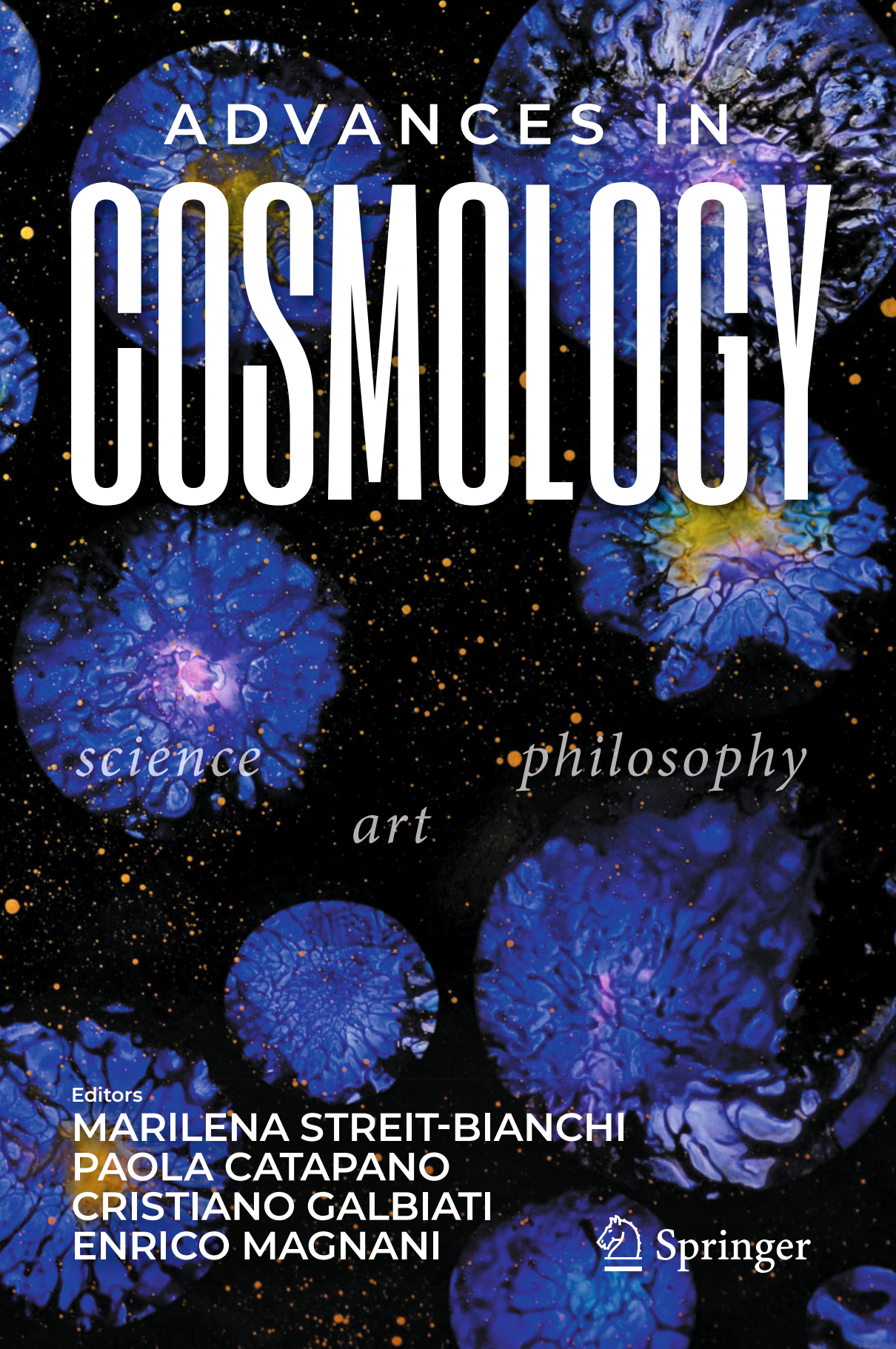
ADVANCES IN
COSMOLOGY
science
philosophy
art
Editors
MARILENA STREIT-BIANCHI
PAOLA CATAPANO
CRISTIANO GALBIATI
ENRICO MAGNANI
Springer

Advances in Cosmology

Marilena Streit-Bianchi · Paola Catapano ·
Cristiano Galbiati · Enrico Magnani
Editors

Advances in Cosmology

Science - Art - Philosophy

Springer

Editors
Marilena Streit-Bianchi
ARSCIENCIA
Wien, Austria

Paola Catapano
CERN
Geneva, Switzerland

Cristiano Galbiati
Princeton University
Princeton, NJ, USA

Gran Sasso Science Institute
L'Aquila, Italy

Enrico Magnani
Reggio Emilia, Italy

ISBN 978-3-031-05624-6 ISBN 978-3-031-05625-3 (eBook)
https://doi.org/10.1007/978-3-031-05625-3

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cover illustration: "Quantum Multiverse n 2" Enrico Magnani 2019, acrylic and enamel on multilayer paper aluminium board, 100 × 76 cm

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Philosophy is written in that great book which ever lies before our eyes—I mean the universe—but we cannot understand it if we do not first learn the language and grasp the symbols, in which it is written. This book is written in the mathematical language.

Galileo, Il Saggiatore (1623)

Foreword

Where do we come from? How did everything begin? How did everything evolve? These questions have interested humankind for thousands of years. People have always been intrigued by the sky and the celestial phenomena they observed, leading them to the science of astronomy, expanding their horizon way beyond the Earth's atmosphere. Later, with the emergence of new technologies and methods, including spectroscopy, a new field emerged: astrophysics. Whereas astronomy deals with celestial objects, space and the physical universe as a whole, astrophysics focusses on the physical processes associated with the objects that form the universe. Both are intimately connected.

According to NASA, the definition of cosmology is "the scientific study of the large scale properties of the Universe as a whole". But you cannot fully understand the large-scale properties without understanding the small scales. Observations of phenomena such as the cosmic microwave background, together with the results of the Standard Model of particle physics, place constraints on the physical conditions that must have prevailed in the early universe. In parallel, phenomena discovered through cosmological observations, such as dark matter and non-zero neutrino mass, suggest the presence of new physics: physics beyond the Standard Model. Answering fundamental questions on our universe, its origin and development, brings researchers and scientists from different fields together. By working together, not only in their respective domains but also across the disciplines, scientists can take great steps in understanding. Collaboration is the key to progress and this book emphasizes this kind of scientific interconnection.

Spreading the net further we should not forget the important role of philosophy. The ancient philosophers were the first to develop an understanding of nature through power of mind alone, described through mathematical principles. We clearly see the interconnectedness of astronomy, astrophysics, astroparticle and particle physics, and philosophy. From the last chapter of Werner Heisenberg's book "Philosophy and Physics", one can conclude that in the course of history it repeatedly turns out that fundamental questions of philosophy related to nature are in reality questions of physics. Philosophy and physics inspire each other, learn from each other.

Art, too, has an important place in humankind's contemplation of the universe. There should be no two cultures: intellectual life should be united in a single culture. Why ? Because reality is very different from what we are able to grasp with our miniscule awareness. We need tools that can take us further. There is as little room in our ability to grasp reality for an infinite Universe with boundaries as there is for a Big Bang. Art, however, sets out to expand our awareness, to create room for new concepts that are just now being researched in science. Or, as Paul Klee once said, art does not reproduce the visible; rather it makes things visible. The same applies to science, especially pure research: it goes far beyond the visible. Sometimes, it is the visions of science that open up new forms to art, and sometimes it is the other way around. Art and science pursue similar visions and I like the idea of Enrico Magnani using the shore as a of merging art with science. Both art and science deviate from the beaten track of thought and perception to conquer something new with great purpose and creativity. Both risk going down the wrong path when wanting to discover new territory. Neither has it easy, because they question the tried and tested and upend the familiar. The mindsets of artists and scientists are very similar. Often they made a choice early in their careers between science and art. Julius von Bismarck, the first artist in residence at CERN, and Enrico Magnani are just two examples. It is art, science and philosophy together that best advance humankind.

Advancement is a complex beast, but it relies on the wellspring of curiosity. We make progress in understanding our world, are exploring others, have put ourselves in a position not only to contemplate the nature of our universe, found exoplanets and made the lives of ordinary people unimaginably better than before. It is fundamental research that forms the basis for all kinds of scientific research. If you had asked Wilhelm Roentgen to invent a tool or a method to show illness or imperfections in the body, he would never have thought of X-rays. Innovation is a virtuous circle linking basic to applied science. Synergy between research and innovation results not only in societal and economic impact but also, and very importantly, in the creation of enhanced opportunities for further developments. This circle, its connection to, and synergies with, other domains of human enquiry needs to remain strong, to be unbroken.

I am sure this book will give you a first glimpse how much the different research areas are interconnected and how much each learns from the others and their respective technologies. It is so important to reach beyond the boundaries of our own research. Doing so brings us forward and helps us to better understand each other, and to better appreciate the universe we share.

Enjoy reading.

February 2022

Rolf-Dieter Heuer
President Council of SESAME,
Director General CERN (2009-2015)
Al Balqa, Jordan

Preface

An unprecedented variety of approaches is available today to researchers whose job is to explore and understand the origin, the structure, the destiny and the laws governing the Universe. This is the underlying reason for this book, whose main motivation is to present Cosmology in its multi-disciplinary aspects as a scientific, philosophical and artistic quest. Since the dawn of time, humans have looked at the sky in an attempt to understand their origins, and the laws governing and influencing them. In most ancient civilizations, astronomers embodied the very power of knowledge. Knowledge was not compartmentalized and their scientific quest often had philosophical implications and visual representations. Natural sciences in those times had no borders. The journey from Astronomy to today's Cosmology has been a long one. Scientists today formulate theories that have to be proven experimentally using the available tools, and this is a very complex process involving the establishment of international collaborations and global networks.

Contemporary cosmology is looking at the Universe as a whole and in the chapters of this book, researchers from all fields explain the ideas, instruments and technologies that are driving the search, what we know and what will be done, through the effort and competence of thousands of experts, to explore and decipher what is still obscure and unknown. Cosmology has been and is the realm of astronomers, physicists and philosophers and, in the past, it inspired religious and mythological cosmogonies.

In 1584, Giordano Bruno, a Dominican friar and philosopher, published the philosophical-cosmological dialogues between Albertino, Burchio, Elpino, Filoteo and Fracastorio in "De l'infinito Universo et Mondi". After demonstrating the infinity of the universe from a logic and theological point of view, Filoteo says "Questo è quel che io dovevo aggiongere. Perché, dopo aver detto l'universo dover essere infinito per la capacità ed attitudine del spacio infinito, e per la possibilità e convenienza dell'essere di innumerabili mondi, come questo; resta ora provarlo... (This is indeed what I have to add; for, having pronounced that the universe must itself be infinite

because for the capacity and aptness of infinite space; on account also of the possibility and convenience of accepting the existence of innumerable worlds like our own; this now still remains to be proven…" (https://www.faculty.umb.edu)).

Scientists are guided in their investigation of Nature by ideas brought by philosophers said in 2013 quantum physicist Francesca Vidotto. The curiosity to understand where we come from and where we are going to is without doubt a strong driving force. What happened during and after the Big Bang, will the accelerated expansion of the Universe continue or the ultimate fate of the Universe is to end in a Big Rip instead of a Big Crunch? These are some of the many questions with philosophical implications raised by modern cosmology.

Many topical researches are at the interface between particle physics, astrophysics and cosmology, offering challenges and opportunities to each of these fields in a common endeavour. Cosmology has been exploring the early and the late universe. Particle physicists have been studying the fundamental elementary particles, the smallest constituents of matter, and their interactions, the properties of antimatter as well as trying to unveil the reasons for the observed matter-antimatter imbalance. Astrophysicists have been studying the nature of astronomical objects and phenomena. However, many still unresolved questions in cosmology are today of common interest. In fact, understanding the origin and nature of dark matter, elucidating the baryon asymmetry, might provide insight into our understanding of the early universe. Will a breaking of the standard Lambda–cold-dark-matter model of cosmology help to elucidate dark matter, dark energy and gravity? Will answer to such questions come from accelerators, telescopes or gravitational waves? Or will the answer come from combined efforts of interacting disciplines? Many are the physicists working to validate or to rule out with experimental data some of the theoretical models proposed, trying to find the model that unifies all forces of nature, describing with a single theory all the phenomena in the Universe, from large to subatomic scale. Thanks to the new technologies, instruments and new probes, as well as their current analysis capability, scientists will provide answers to cosmological questions related to the early universe, thus unveiling new physics. The next decade will surely see theoretical and experimental physicists and astrophysicists analysing and combining the multi-messenger observations collected, thus contributing to a big step forward in our understanding of the early and late Universe, a Universe that they say is 13.7 billion years old. Asking ourselves "Does the Universe have a beginning?" "Do physical laws and causality apply to the universe as a whole?" is asking scientific questions with a philosophical meaning. The discourse about the laws of nature underlying the Universe is expressed in mathematical terms, as Galileo said, and science will continuously push the boundary of human knowledge in the attempt of providing a rational explanation of the world, whereas artists and philosophers will contribute by facilitating our understanding and by stimulating our perception. Clearly, any scientific truth is precarious, as it stays valid until it is not disproved, and this makes scientific knowledge an endless process.

The first section of the book "From Astronomy to Modern Cosmology" is written by renowned theoretical physicists and is divided into three parts, reviewing the past knowledge and assumptions as well as the new theories and ideas, most of them

driving experimental research. Modern cosmology started with the paper "Cosmological Considerations of the General Theory of Relativity" by Einstein, who is considered the Father of Modern Cosmology, and the many observations and discoveries of the twentieth century. Ugo Moschella gives a comprehensive historical overview from Galileo to Einstein. Gabriele Veneziano, who is considered the Father of the String Theory, explains the importance of quantum mechanics for modern cosmology and Kai Schmitz gives an overview of the most modern theories that are driving the contemporary experimental research. Schmitz's research has been recently focussing on the connection between gravitational waves and cosmic strings, with an aim to understand the symmetries and forces that governed the universe during the first moments of its existence.

The second very extensive section, entitled "Dark matter, dark energy, black holes, star formation and other cosmological searches", is presenting how today's experimental research in Cosmology is not the kingdom of a specific area of competence and how the quest for fundamental laws can be and is tackled from different angles, using very different and highly sophisticated technologies, from particle detectors in high-energy accelerators or with detectors placed in space or underground caverns, to telescopes and gravitational wave interferometers. The combined efforts of such large communities of scientists provide remarkable added value in terms of ideas, approaches and competences and are giving a positive push to the cosmological quest.

HEP Physicists at CERN have the tools (particle energy, data analysis capability, etc.) and theories allowing them to search directly or indirectly for Dark Matter at the LHC experiments: ATLAS, CMS and LHCb and, more recently, FASER. Other fixed target experiments (NA64) and others at the Antimatter facility (BASE) are also searching for dark photons, axions or axion-like particles. The search for solar axions started at CERN with CAST, an experiment that started to take data in May 2003 and also, more recently, other experiments aimed at detecting chameleon particles that should be produced when dark energy interacts with the photons; all particles that have not been found yet. The LHC upcoming Run 3 and the High luminosity upgrade are expected to contribute significantly to providing answers to the main unsolved questions of dark matter and dark energy, hopefully showing signatures of some of these elusive particles. After an introduction by Sushita Kulkarni, a theoretical physicist from Graz University leading research on Dark Matter and neutrinos phenomenology with the aim to bridge theories with experimental searches, Caterina Doglioni and Dan Tovey, Deborah Pinna, Carlos Vázquez Sierra and José Francisco Zurita, Marios Maroudas and Kaan Oyboyduman, Michaela Queitsch-Maitland, Dipanwita Banerjee, Stefan Ulmer take us to exploring the tools used in their collaborations to unveil the elusive cosmological signatures that will hopefully let us make the step forward beyond our current understanding.

From CERN, the journey continues on board of the ISS (International Space Station) with the chapter "The AMS experiment on the International Space Station", by Maura Graziani and Nicola Tomassetti. They tell us how the observation of 194 billion cosmic ray events recorded from the start of the operation of the detector (May 2011) to November 23, 2020 in a collaboration of 44 institutions from the

USA, Europe and Asia has provided key information on the origin of cosmic ray electrons, the propagation of cosmic particles and how they are using this knowledge to understand cosmic dark matter and antimatter. Claudio Bortolin and Paola Catapano, in their "The right key- Four spacewalks to repair the Alpha Magnetic Spectrometer on the International Space Station", tell us about the challenges of the spacewalks conducted by Italian Astronaut Luca Parmitano (ESA) with his US colleague Drew Morgan (NASA) for the crucial repair which enabled the full functionality of AMS-2. The detector is expected to take data until the ISS remains in orbit.

"Under the Gran Sasso" by Cristiano Galbiati and Walter Bonivento takes us to the underground area beneath the Gran Sasso mountain and the Laga National Park in Italy, where astrophysics research for WIMPs (Weakly Interacting Massive Particles), a Dark Matter candidate, using the Argon (DarkSide-20k) or Xenon (XENONnT) detectors is conducted. They tell us about their choice to use Argon and the technologies developed to reach the needed zero background. The collaboration says it is not axions or physics beyond standard model.

"Gravitational waves: Why and How" by Federico Ferrini, former Director of EGO (European Gravitational Observatory) at the time when the first gravitational wave signals were observed, introduces non-specialist to the fascination of the gravitational universe and explains how gravitational wave cosmology started and developed. Interesting to note that since the first detection in September 2015 to November 2021, the LIGO, Virgo and KAGRA interferometers recorded 90 events or cosmic quakes (79 of them have been recorded between April 2019 and March 2020), generated by the merging of black holes, neutron stars or even of a neutron star with a black hole. The analysis of the information provided by these events is expected to give new clues on how stars live and die. Interferometers are today key cosmological probes to understand the gravitational Universe.

Looking at the sky with a telescope has been the way in which astrophysicists have explored stars, planets, galaxies and our solar system since the dawn of science. The chapter by Gianni Marconi and Riccardo Scarpa entitled "Big telescopes and observatories: hi-tech challenges for great astronomical science" gives an insight into optical astronomy and the work of astrophysicists today. As Massimo Tarenghi, Astrophysicist and former Director of VLT and ALMA, once said *In the old times, I was looking through an eyepiece for hours, had to make sure that the telescope was not going away from a guiding star and the record was done using a photographic plate. What a change to today's modern instruments!* To make precise astronomical instruments requires considerable human toll, the effort of large communities of researchers and a crucial follow-up with industrial partners. Many high technological developments are involved in the production of mirrors and the infrastructures allowing very precise movement of these gigantic structures. Just as an example, the European Extremely Large Telescope (E-ELT) has a 39-m-diameter mirror. Several scientific communities are using the telescope at different times looking at the origin of signals farther out in space and further back in time. The bigger revolution in the astronomical science, at the same time as that of the instruments, regards the way in

which data is collected and processed, thanks to the creation of big archives accessible to astrophysicists that analyse the data and use them in a variety of ways and for a variety of purposes. Telescopes scan the sky recording tiny changes; new sophisticated telescope platforms measure the strength of gravitational waves produced during the epoch of inflation 380.000 years after the Big Bang. Cosmic-microwave background (CMB) measurements in multiple frequency bands allow the removal of the galactic foreground. The JWST telescope launched on 25 December 2021 is showing sharp images and has already discovered a candidate new galaxy, the Maisie's galaxy, so far the most ancient ever observed 300 Myr after the Big Bang.

JWST is opening new revolutionary avenues to understanding the origin of our Universe and the early evolution of galaxies as well trying to elucidate the existence of dark matter.

In chapter "Other worlds in the cosmos: from philosophy to scientific reality", Michel Mayor, Nobel prize laureate for physics in 2019, with Jim Peebles and Didier Queloz, for the discovery of exoplanet *51 Pegasi b* orbiting a solar star, with other young professors from the University of Geneva, Emeline Bolmont, Vincent Bourrier, David Ehnreich and Bern, Christoph Mordasini are bringing us to discover the habitability, the upper atmosphere and the atmosphere of the exoplanets and their formation. The existence of exoplanets originated from a philosophical assumption before it was scientifically proven. Today, the analysis of planetary atmospheres may reveal spectral characteristics induced by the development of life; advances in spectroscopy studies of exoplanets, as the authors say, make the search for extraterrestrial life possible. According to NASA, 4000 exoplanets have been discovered and confirmed so far and about 1000 more candidates are on the waiting list.

The third section, on Philosophy, focusses on the link existing between science and philosophy. In the past and until the eighteenth century, there was no dividing line between the scientific disciplines and philosophical considerations were often associated with scientific observations; some representations to explain the advancements in natural sciences as well as by astronomers can be seen as pieces of art. It is worth to remember that at CERN and elsewhere the Bell theorem with the locality and causality assumptions and the Bell's inequalities have been debated not only for their scientific but also for their philosophical implications. Today, Cosmological philosophy starts to be a domain by itself. Francesca Vidotto, a theoretical physicist whose research is focussed on Loop Quantum Gravity and, in particular, on spinfoam Cosmology, explains in the chapter "Space, time and matter in the primordial Universe" that formulating a new theory of understanding brings scientists to the edge of knowledge where philosophy and experience must meet.

The fourth and final section is dedicated to Art. Enrico Magnani, a nuclear engineer by training and currently a full time artist, presents his artwork, which recently focussed on Supernovae, Dark Matter and Multiverses, in a chapter entitled "The shore between Art and Science". The chapter tries to answer questions such as how much technology is affecting the artistic work, how much scientific ideas are nurturing art and how much scientists have been inspired or contributed to art and philosophy. It also identifies some common traits in artistic and scientific research: the quest to unveil the mystery, to push the boundaries of human knowledge more and

more into the unknown, using beauty and intuition as a compass. Science, Philosophy and Art are expressions of a common quest: the necessity we pursue since the dawn of time, to understand the world and the universe we are living in and of which we are an integral part.

The divide between cultures has been a long and debated issue since Charles Snow invoked the gap between science and humanities. The present book not only explains where we stand today in Cosmology and what are the main questions and the tools and ideas to provide scientific answers, but it also establishes a dialogue between cultures, based on knowledge, in an attempt at crossing the existing boundaries and making the gap less deep.

We do wish you happy reading.

Wien, Austria — Marilena Streit-Bianchi
Geneva, Switzerland — Paola Catapano
February 2022

Acknowledgements

The editors would like to warmly thank all the authors who have embraced this project and contributed with their valuable texts and illustrations to make this book interesting and attractive. We also thank *Springer Nature* and, in particular, *Springer's* Executive Editor Marina Forlizzi for her valuable support and assistance in the making of this book. The amount of ideas and research going on in the many sectors of physics and astronomy and the technological developments that made the most recent research in Cosmology possible are presented. The book outlines how today multi-disciplinary approaches are fruitfully contributing to the search that is trying to bring out the dark from the unknown. The knowledge and assumptions cosmologists are making today, both at theoretical and experimental levels, in tackling the origin and development of the universe, and how this human questioning is also pervading philosophy and art will accompany the reader through a long journey of discovery.

In most ancient civilizations, astronomers incarnated the power of knowledge. We intended, by connecting Science with Philosophy and Art, to remind people that many of the questions science asks at fundamental levels are a quest of our humanity. To conclude, using the words of one of the authors: R*esearch, Beauty, Intuition, Curiosity are the elements that we find on the shore between art and science.*

Contents

About the Editors

Marilena Streit-Bianchi was born in Rome (Italy). She received a Doctorate in Biological Sciences from the University of Rome and joined CERN, the European Organization for Nuclear Research in Geneva (Switzerland), in 1969. She has been a pioneer in the study of high-energy particles produced by accelerators for cancer treatment. She has held managerial positions on safety training and technology transfer, has been a senior honorary staff member at CERN and is actively engaged in art and science as a book editor and curator of exhibitions in Europe and Mozambique. She is the Vice President of the International Association ARSCIENCIA and member of the Italian Physical Society (SIF). She has been one of the editors of the book "Mare Plasticum-The Plastic Sea. Combatting Plastic Pollution Through Science and Art".

Paola Catapano was born in Lucera (Italy). She is a Science Journalist and Science Communicator at CERN. In 1987, she graduated in simultaneous interpreting from the University of Trieste and in 1997 she obtained a master's degree in Science Journalism from ISAS (International School for Advanced Studies), Trieste. At CERN since 1990, she has covered several positions from assistant to the Director General Carlo Rubbia, to leader of public outreach. She is currently head of the editorial content production in CERN's Education, Communication and Outreach group. In 2010–11, she authored and hosted the TV program DIXIT Scienza for the Italian National Television, RAI. She created and led the research expedition to the Arctic Polarquest 2018, whose program included historical exploration, measurements of cosmic rays, drone mapping of uncharted territories and plastic debris sampling until the 82°nd parallel (cf. and book "Mare Plasticum-The Plastic Sea-Combatting Plastic Pollution Through Science and Art"). She has participated in several Science Festivals and won international awards and is the author of the book *Il lungo viaggio delle onde gravitazionali*, Textus, 2021.

Cristiano Galbiati was born in Milano (Italy). He is a Particle Physicist with a Ph.D. from the University of Milan. He is Full Professor at the Physics Department of Princeton University (Princeton, New Jersey) and Professor of Particle Astrophysics at the Gran Sasso Science Institute (L'Aquila, Italy). His research has been on solar

neutrinos and is at present Coordinator of the international experiment DarkSide at the INFN Gran Sasso National Laboratory and of the project Aria at Monte Sinni in Carbosulcis (Gonnesa, Sardinia), searching for dark matter. Besides his many scientific publications, he is the author of the book *Le entità oscure. Viaggio ai limiti dell'Universo,* Feltrinelli, 2020.

Enrico Magnani was born in the province of Reggio Emilia (Italy). He is an internationally recognized artist whose work integrates art, science and transcendence. He graduated as nuclear engineer at the Polytechnic University of Milan in 2004 and worked as a Researcher at the Karlsruhe Institute of Technology (KIT) for the nuclear fusion projects ITER and DEMO (2006–2010). He has been painting and doing artworks since his young age and at a certain point in his scientific career, feeling that knowledge only based on rational speculations was not fulfilling his life, he decided to dedicate himself entirely to art. After the first figurative period (1995–2006), his work became more and more abstract, using various kinds of materials and techniques. His work has been presented in museums, foundations, private galleries and public institutions in Europe and the United States. In 2017, he started a series called "Supernova" and in July 2019 he exhibited at CERN his "Searching the Unknown—The Dark Matter Collection". In autumn 2020, he completed his permanent installation "Quintessence", now visible at the INFN Gran Sasso National Laboratory in L'Aquila. He holds seminars and training workshops on creativity.

From Astronomy to Modern Cosmology

Looking at the Sky: From Pythagoras to Einstein Through Galileo and Newton

No Admission Without Knowledge of Geometry

Ugo Moschella

The material world—the reality—is not something given, but is born with us. For the "given" to become reality, it must be resurrected in the literal sense of the word. This is the role of Science, this is the role of Art.

[a]Ossip Mandelstam, Letter to Marietta Shaginyan, April 5, 1933.

1 Prelude

E quando miro in cielo arder le stelle;
Dico fra me pensando:
A che tante facelle?
Che fa l'aria infinita, e quel profondo
Infinito Seren? che vuol dir questa
Solitudine immensa? ed io che sono?
…
E dell'innumerabile famiglia;
Poi di tanto adoprar, di tanti moti
D'ogni celeste, ogni terrena cosa,
Girando senza posa,
Per tornar sempre là donde son mosse;

U. Moschella (✉)
DISAT—Università dell'Insubria, Via Valleggio 11, Como, Italia
e-mail: ugo.moschella@uninsubria.it

INFN, Sez. di Milano, Via Celoria 16, Milano, Italia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_1

Uso alcuno, alcun frutto
Indovinar non so.[1]

Giacomo Leopardi's wandering shepherd looks up into the sky, searching for the meaning of all things (*A che tante facelle?*) and of its very life (*ed io che sono?*). The attempt to find in the skies the point and the purpose of all the nowhere-going movements and struggles of everything and every human being (*di tanto adoprar, di tanti moti*) is doomed to failure (*Uso alcuno, alcun frutto / Indovinar non so*).

The questions raised by the shepherd have been asked by men and women of all epochs, though rarely expressed with such intense words. The gesture of looking at the starry sky seems to be specific to the genus homo as much as the *nosce te ipsum* [1] to such an extent that it could be also taken as a definition of what being human is all about.

The terrestrial landscape has changed a great deal, but, to the naked eye, the sky has not changed at all, because the couple of million years elapsed after the appearance of the first *homines* on Earth are, on a cosmic scale, *like a day that has just gone by, or like a watch in the night*.[2]

All men and all women have seen the same sky. Yet, their cosmogonic narrations and their cosmological representations are very different from each other and above all are radically different from that which is supposed to be the modern scientific vision of the world, our vision. The story of the ideas that led to this vision is extraordinarily fascinating, dramatic and sometimes tragic and has been narrated a thousand times in books that have rightly become classical. We will go over some of its highlights again [2].

2 A Name, an Idea

The idea of universe is not a primitive idea [3]. If, as customary, we let history begin with the invention of writing around the year 3000 BC, we see that humanity was able to do without the idea of universe, or, better to say, its explicit thematization, during half of its history, not to mention the immensity of prehistory: a word to designate the totality in a unified way appeared in Greece only around the year 500 BC.

Previously, a more or less exhaustive enumeration of the things contained in the totality or else a binary opposition were used—the biblical and Homeric formula *Heaven and Earth* being the best known. It is only when this distinction between the things on which we can—in principle—have an influence and those which are

[1] *…And when I gaze upon the stars at night – In thought I ask myself – "Why all these torches bright? – What mean these depths of air, – This vast, this silent sky, – This nightly solitude? And what am I?" – …– "And all this mighty motion, and this stir – Of things above, and things below, – No rest that ever know, – But as they still revolve, must still return – Unto the place from which they came, – Of this, alas, I find nor end nor aim!"*
Giacomo Leopardi, Canto notturno di un pastore errante dell'Asia (excerpt) 1830.

[2] Psalms 90, 4.

completely beyond us is put aside, that the "world" can appear. Wittgenstein expresses this state of affairs with the formula "The subject does not belong to the world, but it is a frontier of the world".

Legend has it, that was Pythagoras to choose the proper noun: "cosmos", which, as everyone knows, opposes itself to chaos and designates order and beauty, or to say it better, the beauty that derives from the order. "Pythagoras was the first, who named the encompass of the whole a Cosmos, because of the order which is in it" [4].

The Latin name "mundus" has exactly the same meaning as cosmos.[3] Pliny the Elder tells us in his *Naturalis Historia* that "The Greeks gave to all things the name 'cosmos' and we called it 'mundus' by virtue of its perfect and absolute elegance." The name "universe" (*Unvorsum*), a poetic contraction of *unus* and *versus*, appears for the first time in the fourth book of Lucretius' *De Rerum Natura*. Lucretius gives this word the meaning of a set of particles that rotate all together.

The thirtieth fragment of Heraclitus gives us a glimpse of the eternal cosmic order, that of a self-sufficient totality which does not require external instances: "This world, which is the same for all, no one of gods or men has made; but it was ever, is now, and ever shall be an ever-living fire, with measures of it kindling, and measures going out."

On the contrary, according to Plato, order does not pre-exist the primordial chaos of the chora but it results from the creative action of a demiurge. Timaeus recounts the mythical birth of the universe: "The world …has become a visible living creature containing the visible—the sensible God who is the image of the intellectual, the greatest, best, fairest, most perfect—the one only begotten heaven."

The order of the universe is not only the visible manifestation of the intelligible God; it is also the model to be imitated to return to the original state of excellence, which was lost by the incarnation of the soul. Cosmology will keep this ethical dimension for two millennia, until the birth of the scientific vision of the world. Thus, the word cosmos–order is already a "cosmology". It gives a description of the totality that is not neutral but implies a judgment of value. Perhaps, it is interesting to compare this stance with the modern point of view, exemplified here again by Wittgenstein's words: "The sense of the world must lie outside the world. In the world everything is as it is and happens as it does happen. In it here is no value—and if there were, it would be of no value." As for the means to describe and try to understand the order of the cosmos, the Greeks also explain to us the relative roles of "physics" and "mathematics": "The task of the contemplation of nature (*theoria phusikè*[4]) is to examine the substance of the sky and the stars, the power and the quality of generation and corruption, and, by Zeus!, it is capable of leading demonstrations on the subject of the size the form and the order of things. As for astronomy (*astrologia*) it does

[3] The original sense of "woman's ornament" is metaphorically turned into order and beauty, the beauty resulting from order (cosmetics). The usage was for a long time perceived as a metaphor. The "cosmic" unique usage took centuries to emerge. Mundus is the etymological source of the Italian word "monile".

[4] To perceive the *logoi* in beings is the act known as *theoria phusikè*, the second of the three stages of the spiritual life distinguished by Evagrius and the tradition that followed him [5].

not undertake to speak of anything like that, but it demonstrates the order (*taxis*) of celestial things, having declared that the sky (*ouranos*) is truly a cosmos; it speaks of forms, sizes, distances from the Earth to the Sun and the Moon, eclipses, conjunctions of stars, on the quality and quantity that are shown in their revolutions." (Posidonios, 135-51 avant J.-C.). The *theoria phusikè* has therefore the task of examining the substance of the sky and the stars. On the other hand, mathematics must be limited to saving the appearances. This warning will return dramatically seventeen centuries later.

3 Aristotle and the Ptolemaic World

Physics enters in cosmology with Aristotle: the physical foundations of the standard cosmological model of the ancient world are rooted in the Aristotelian conceptions of movement and gravity. The Stagirite distinguishes between three types of movements. Two of them occur in the sublunary world: the natural movements of falling heavy bodies (made in prevalence of earth and water) and of rising light bodies (made in prevalence of air and fire) are caused by their tendency to proceed to their "natural place"; on the contrary, violent movements require an external force as a cause.

The very existence of a natural place explains the central position and the spherical shape of the Earth. It also explains what gravity is. An apple falls because it aims to go where heavy bodies naturally go. That place is necessarily at the center of the universe, where the Earth is located (otherwise it would also end up falling there). Moreover, the Earth cannot spin around its axis nor can it revolve around the Sun because the perfect circular movement cannot exist in the changeable and corruptible sublunary world. Terrestrial creatures move on straight and irregular trajectories, because they are limited and imperfect and must seek food and help outside of themselves.

The motionless center of the cosmos is therefore not a place of delight, but rather a garbage dump where all the heaviness of the sublunary world falls. And yet it is the unique and privileged center around which the spheres of the superlunary world revolve, bringing the stars with them in their race without beginning or end.

The heavens are concentric crystalline spheres made of the fifth element: the aether or quintessence. The aether has no weight or lightness and therefore cannot go towards the center or away from it: its movement is by nature circular and uniform. Eudoxus of Cnidus, a disciple of Plato, had invented them as a calculation device but Aristotle considers the spheres of the heavens as physically existing. There are fifty-five of them and the last one is fixed and borders the finite universe; indeed, if the universe has a center, it can only be finite.

Can we bend outside this last frontier? The question makes no sense because there is no outside. There is nothing. Not even the void…"It is evident not only that there is not, but also that there could never come to be, any bodily mass whatever outside the heavens. …There is also no place or void or time outside the heaven. For in every place body can be present; and void is said to be that in which the presence of body, though not actual, is possible; and time is the number of movement. But

in the absence of natural body there is no movement, and outside the heaven, as we have shown, body neither exists nor can come to exist." [6].

The world has to wait for Giordano Bruno to meet the one "who has pierced the air, penetrated the sky, toured the realm of stars, traversed the boundaries of the world, dissipated the fictitious walls of the first, eighth, ninth, tenth spheres, and whatever else might have been attached to these by the devices of vain mathematicians and by the blind vision of popular philosophers." [7].

Perfected by Ptolemy in the Almagest and in the Hypotheses planetarum, the system of the spheres (and epicycles) has been the foundation of the standard vision of the world for centuries. It accounts for the celestial movements of the stars with good precision. It also gives a cosmological basis to anthropology and ethics, extending the "scientific" representation into an answer to the question about being-in-the-world. That world would collapse under the deadly blows of the *De Revolutionibus Orbium Coelestium* by Nicolaus Copernicus.

4 The Copernican Revolution

In fact, canon Copernicus was not a revolutionary. His inspiration and his cosmological principles were strongly linked to the traits of Aristotelianism described above: they are the perfection of circular movements but also the finiteness and the spherical shape of the universe and the solidity of the crystalline spheres. And more than observing the sky, Copernicus, as a good humanist, sought his sources in the classics: "…I began to be annoyed that the movements of the world machine, created for our sake by the best and most systematic Artisan of all, were not understood with greater certainty by the philosophers, who otherwise examined so precisely the most insignificant trifles of this world. For this reason I undertook the task of rereading the works of all the philosophers which I could obtain to learn whether anyone had ever proposed other motions of the universe's spheres than those expounded by the teachers of astronomy in the schools. And in fact, first I found in Cicero that Hicetas supposed the earth to move. Later I also discovered in Plutarch that certain others were of this opinion." [8]. In fact, the Copernican revolution which "places the Earth as mobile and the Sun, on the other hand, as immobile at the center of the universe, is based on exactly the same astronomical data of the *Almagestus*.[5] There was nothing new under the Sun (nor above). Except that once the Earth is removed from the center of the universe, a question that we thought was decided comes back strongly: what is gravity?

[5] And also on the astronomical data transmitted by the Arabs. Albategnius and some other Arab astronomers are quoted by Copernicus.

Until recently it was believed that the *De revolutionibus* was already completed in 1530. Today it is known that the sixth book was written only after 1539. Four hundred copies were printed only in 1543, shortly before the author's death. The initial print run was not sold out. The text is preceded by a preface written anonymously by Andreas Osiander who had been commissioned by Georg Rheticus, the author of the *Narratio prima* [9], to oversee the publication of the book. Osiander was a former Catholic priest turned Lutheran theologian, very active and vaguely heretical. By professional deformation, he saw rather well the risks inherent in the theses of Copernicus, theses that undermined the scientific bases of the cosmic order that philosophy and theology conceived as anthropocentric. To counter these risks, Osiander, in his anonymous preface *To the Reader Concerning the Hypothesis of This Work*, repeats in even more drastic terms Posidonius' arguments:

"Since the novelty of the hypothesis of this work has already been widely reported, I have no doubt that some learned men have taken serious offence because the book declares that the earth moves, and that the sun is at rest in the center of the universe; these men undoubtedly believe that the liberal arts, established long ago upon a correct basis, should not be thrown into confusion. But if they are willing to examine the matter closely, they will find that the author of this work has done nothing blameworthy. For it is the duty of an astronomer to compose the history of the celestial motions through careful and expert study. Then he must conceive and devise the causes of these motions or hypotheses about them. Since he cannot in any way attain the true causes, he will adopt whatever suppositions enable the motions to be computed correctly from the principles of geometry for the future as well as the past. The present author has performed both these duties excellently. For these hypotheses need not to be true nor even probable. On the contrary, if they provide a calculus consistent with the observations that is enough." [10].

A computational hypothesis, which concerns only mathematicians, that's all. Revolution is something different!

Yet, the revolution was secretly underway. On the evening of November 11, 1572, leaving his uncle's underground alchemical laboratory and looking towards the zenith, Tycho Brahe, the greatest ever observer of the sky with the naked eye, saw a "nova et nullius ævi memoria prius visa Stella," a new star, brighter than Venus, in the constellation of Cassiopeia. It was unheard of! He doubted his vision and asked the peasants who were passing by if they saw the same star as him in the sky. This event would change the life of Tycho who became the first of the modern astronomers. The Stella nova was there to destroy the idea of immutability of the heavens, because there were changes in the superlunary world. Stars could be born and perhaps die …

After that, it was the turn of the crystalline spheres to break apart under the blows of the great comet of 1577. Tycho observed it for several months; the parallax of the comet allowed him to decide a thousand-year-old question: the comets were indeed celestial bodies. "All the comets which I have observed move in the ethereal region of the world and never in the sublunary region as Aristotle and his followers wanted us to believe for many centuries!" And as its trajectory, which was not at all circular, went through the orbs of the planets "the reality of crystalline spheres must be excluded from the heavens." The spheres do not really exist, the sky is free,

open in all directions and there is no obstacle to the race of the planets. But, once destroyed the crystalline spheres *a quo moventur planetae*? What is the cause of the motion of planets and other celestial bodies?

Finally, the last to dissolve was the circular motion of the planets, already undermined by the superlunary comets. It is all the more ironic that the main motivation of Copernicus' work was to reestablish the perfection of the circular and spherical geometry. However, the astronomical data that Tycho had entrusted to Johannes Kepler and that Kepler had been studying hard for six years, said something else: the orbit of the planet Mars was not circular, nor reducible to a composition of circles, but it was an ellipse with the Sun at one of its focuses. Kepler wrote the *Astronomia Nova,* αιτιολοητος *seu physica coelestis, tradita commentariis de motibus stellae Martis ex observationibus G.V. Tychonis Brahe*, a book that since 1609 marks forever the history of astronomy. Yet the planetary ellipses were to remain a dead letter for a long time. It is only after Newton's law of universal gravitation that everybody accepted the Keplerian orbits of the planets. Today they still remain elliptic (roughly!)

5 The Galileo Affair

The *Astronomia Nova* was unlucky to appear shortly before the publication of Galileo's *Sidereus Nuncius* (1610). The Galilean heavenly messenger announces the revolution in broad daylight. He reveals that the Copernican point of view is not just a technical question for mathematicians but concerns everyone. Since then, the Galileo affair has been one of the most significant events in the history of western culture [11, 12]. Countless literary, philosophical and scientific books have supported virtually every possible stance regarding Galileo's condemnation, the relationship between science and religion, the birth of modern science and the scientific method. We are going to mention here only the few aspects that are relevant for our cosmological tale.

Galileo propagated the revolution by means of a canon-shaped telescope—the *cannocchiale*. This Dutch-made instrument was originally an object for the amusement of the wealthy. Galileo perfected it obtaining a much better magnification and sold it to the Senate of Venice as a military instrument.

Then, on the evening of August 25, 1609, he pointed his perfected telescope towards the sky and discovered a world that no one had ever seen. He observed the lunar landscape with mountains and valleys and myriads of stars of the Milky Way. On January 7, 1610, he observed three stars near Jupiter and then, on January 13, a fourth one. Their positions had changed: they were circling around Jupiter as everyone could see. Freed from verbose debates ("e noi liberati da verbose discussioni") the millenary conception of an unchanging and perfect sky that revolves around the Earth was over. Six and a half years later, on March 5, 1616, Copernicus' *De Revolutionibus orbium coelestium* was added to the Index, sixty-three years after its publication.

All of this is well known. But there is something here that must not escape our post-modern eyes. Today, Galileo's gesture of observing the sky with his *cannoc-*

chiale may seem obvious and even obsolete, but it was not at all so in his epoch and for many reasons. First there was nothing to see in the sky. Everything had to go as it had always been. Worse, the instrument used to observe the sky was unworthy, made by mechanicians and engineers, and therefore not very commendable for honest gentlemen and for academics (who often stubbornly refused even to touch the telescope).

It is by disregarding this official science that, with his solitary gesture, Galileo abandons the conception of the human natural senses as an absolute criterion of knowledge, and, trusting in what he sees through his instrument, lays the foundations of the scientific revolution that has forever changed the history of humanity.

Copernicanism is also the pillar of the project to establish a new science. The Copernican overturn of the cosmic order leads to the revolutionary idea that there is only one physics that governs the movements on Earth as in Heaven and opens the way for geometry to come down in our sublunary world to explain celestial and terrestrial phenomena on the same basis. Heaven, so to speak, descends to Earth (Fig. 1).

The most recent Galilean studies indeed point out how erroneous it would be to separate Galileo's researches on the movement and the fall of massive bodies—made in Pisa from the 1580s and in Padua thereafter—from his subsequent astronomical studies. These studies had already led him to reject scholastic physics. Aristotle taught, for example, that a body weighing ten pounds falls from a certain height in ten times less time than a body weighing one pound. Legend has it, that the young Galileo climbed the tower of Pisa and, at the passage of the academic procession, dropped the two weights that arrived on the ground almost at the same time. Experimental evidence had never been sought for before. It was an absolute novelty of the new Galilean scientific method.

In the bitterness of his old age, the prisoner Galileo had nevertheless the courage to go back to the study of free fall and the movement of the projectiles, in the third and fourth parts of his last book, *Discorsi e dimostrazioni matematiche intorno a due nuove scienze* (1638). "La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi (io dico l'universo), ma non si può intendere se prima non s'impara a intender la lingua, e conoscer i caratteri, ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezi è impossible a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto."[6] [13].

[6] Philosophy is written in this grand book, which stands continually open before our eyes (I say the 'Universe'), but can not be understood without first learning to comprehend the language and know the characters as it is written. It is written in mathematical language, and its characters are triangles, circles and other geometric figures, without which it is impossible to humanly understand a word; without these one is wandering in a dark labyrinth.

Fig. 1 Albrecht Dürer: *Melancholia I* (1514). Geometry comes down to Earth from the Skies. The wings of the angels are now useless…[14]

To better appreciate the greatness of this last work, we must remember that in the seventeenth century geometry had nothing to do with physics in the sublunary world (Fig. 1). Even the eminently practical problem of calculating the trajectory of a projectile posed by the new techniques of artillery, was approached by compulsorily studying the Physics of Aristotle to find the correct way to compose violent movements with natural movements (Fig. 2). In this context, it is possible to understand the novelty and importance of the experimental works of the young Galileo and of the Discorsi of his maturity where he shows that the trajectory of a projectile is a parabola resulting from the composition of two movements that do not interfere with each other: a straight horizontal motion in accordance with the principle of inertia and a uniformly accelerated vertical motion whose acceleration does not depend on the mass of the body.

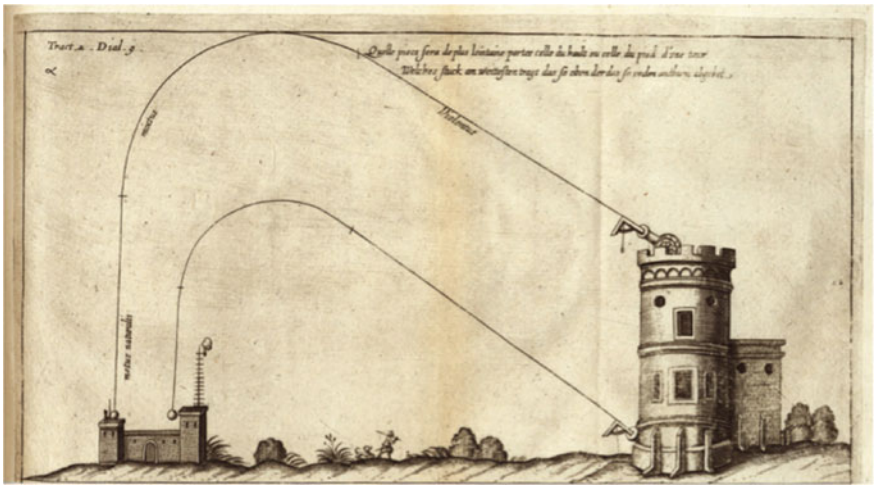

Fig. 2 Diego Ufano: *Artillerie (1628)*

The principle of equivalence of Galileo stating that all bodies fall (in void) with the same acceleration is, perhaps, the most important result of Galileo's new experimental science. It contains the germ of an answer to the questions about the nature of gravity and its universality (it is universal because it acts in the same way on all things and because it founds every science of the universe—including Aristotle's cosmology!) But this (provisionally) definitive answer will not come until three centuries later. Meanwhile Galilean science will find its climax and its accomplishment in the work of the greatest man of science ever: Isaac Newton (Fig. 3).

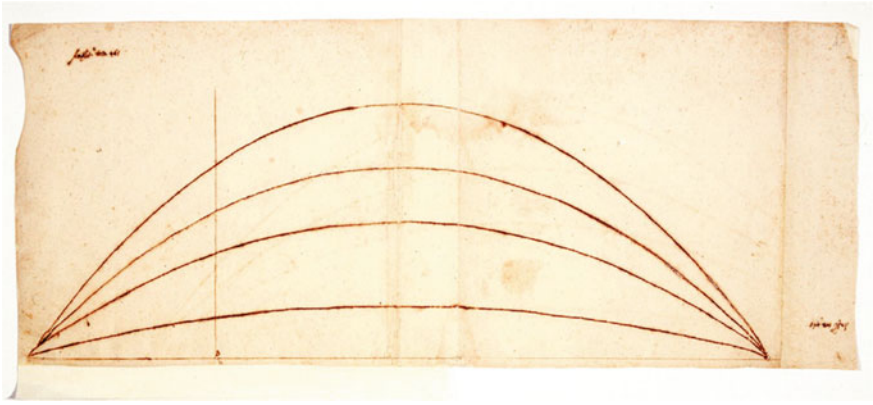

Fig. 3 Galileo Galilei. Codex 72. Folio 42 r. Galileo's new geometric theory of movement applied to ballistics

6 Never at Rest: Newton

It is with Newton that the discussion on the nature of space and time bursts into the *Philosophia Naturalis*, that is to say in the physical science. The first book of his immortal work, the *Philosophiae naturalis principia mathematica*, starts with the *Scholium* focusing on the notions of the three elements that make up the universe, space, matter and movement: "I do not define time, space, place, and motion, as being well known to all. Only I must observe, that the common people conceive those quantities under no other notions but from the relation they bear to sensible objects. And thence arise certain prejudices, for the removing of which it will be convenient to distinguish them into absolute and relative, true and apparent, mathematical and common. […] Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external, and by another name is called duration. Relative, apparent, and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time; such as an hour, a day, a month, a year. Absolute space, in its own nature, without relation to anything external, remains always similar and immovable. Relative space is some movable dimension or measure of the absolute spaces; which our senses determine by its position to bodies; and which is commonly taken for immovable space." [15].

The almost insurmountable problem is that there is no hold on absolute space. One can even doubt its existence, as will do, among others, Huygens, Leibniz and Mach. Newton responds to this objection with the famous example of a vase that contains water and they rotate together: the concavity of the free surface of water is the proof of the movement of water "relatively" to absolute space. It is once again the circular movement that plays a distinguished role; here it offers us the possibility of discerning absolute movements from relative movements. But doubts remain, and it is precisely by resuming this ideal dialogue with Newton on absolute space and the relativity of inertia that Einstein, in 1917, will lay the foundation of the modern scientific cosmology.

Newton speaks about inertia and the universality of gravity in the third book of the *Principia*. The crucial focus is given in the third *Regula Philosophandi* at the beginning of the book: "All bodies are moveable, and endowed with certain powers (which we call the *vires inertiae*) of persevering in their motion, or in their rest we only infer from the like properties observed in the bodies which we have seen. […] Lastly, if it universally appears, by experiments and astronomical observations, that all bodies about the earth gravitate towards the earth, and that in proportion to the quantity of matter which they severally contain, that the moon likewise, according to the quantity of its matter, gravitates towards the earth; that, on the other hand, our sea gravitates towards the moon; and all the planets mutually one towards another; and the comets in like manner towards the sun; we must, in consequence of this rule, universally allow that all bodies whatsoever are endowed with a principle of mutual gravitation. For the argument from the appearances concludes with more force for the universal gravitation of all bodies than that for their impenetrability; of which,

among those in the celestial regions, we have no experiments, nor any manner of observation. Not that I affirm gravity to be essential to bodies: by their vis insita I mean nothing but their *vis inertiae*. This is immutable. Their gravity is diminished as they recede from the earth."

Newton then sets out the law of Universal Gravitation: two bodies attract each other with a force proportional to the product of their masses and inversely proportional to the square of their distance. This law applies to the apple that falls on the Earth (and the Earth that falls on the apple) as to the Moon that falls around the Earth and to the planets that turn around the Sun. Newton accomplished Galileo's project of unification and abolished the principle of a substantial difference between Heaven and Earth. Better still: a single law accounts for a wide variety of phenomena. Newton's law is now part of the culture of teenagers from all over the world (those who have the chance to go to school).

And yet—as everyone knows—Newton declares himself ignorant as to the physical reality of gravitational attraction: "Hitherto we have explained the phenomena of the heavens and of our sea, by the power of Gravity, but have not yet assigned the cause of this power. […] I have not been able to discover the cause of those properties of gravity from phenomena, and I frame no hypotheses. For whatever is not deduced from the phenomena, is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy."

Newton had tried to find an explanation for gravity as a contact force caused by invisible particles, but he realized that this explanation could not work. In a letter to Richard Bentley he wrote: "You sometimes speak of gravity as essential and inherent to matter. Pray do not ascribe that notion to me, for the cause of gravity is what I do not pretend to know and therefore would take more time to consider of it. It is inconceivable that inanimate brute matter should, without mediation of something else which is not material, operate upon and affect other matter without mutual contact, as it must be if gravitation, in the sense of Epicurus, be essential and inherent in it. And this is one reason why I desired you would not ascribe innate gravity to me. That gravity should be innate, inherent, and essential to matter, so that one body may act upon another at a distance through a vacuum, without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it."

After such a long journey, the question *what is gravity* remains unanswered.

The *Principia* do not contain cosmological assertions *per se*. The cosmological question is briefly mentioned in the correspondence with Richard Bentley. One of Bentley's questions is still relevant to those seeking to understand the formation of cosmic structures, and can be summarized as follows: a uniform distribution of matter in space may give birth to the Universe as we see it by purely natural causes? Newton replied that an infinite universe was needed: "As to your first query, it seems to me that if the matter of our sun and planets and all the matter in the universe were evenly scattered throughout all the heavens, and every particle had an innate gravity

toward all the rest, and the whole space throughout which this matter was scattered was but finite, the matter on the outside of the space would, by its gravity, tend toward all the matter on the inside, and by consequence, fall down into the middle of the whole space and there compose one great spherical mass. But if the matter was evenly disposed throughout an infinite space, it could never convene into one mass; but some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered at great distances from one to another throughout all that infinite space."

The idea of an infinite universe was not new. Giordano Bruno had already advanced it more than a century before [7]: "The universe is infinite and therefore there is no body in it to which it would belong to be at the center or on the periphery or between these two extremes." And why would the universe be infinite according to Bruno? Because there is no reason why it shouldn't be. Bruno appeals to the principle of sufficient reason[7] a century ahead of Leibniz and affirms the primacy of intellectual knowledge over sensible knowledge [16]: "I am sure that it will never be possible to find even a half-probable reason why there should be a limit to this corporeal universe, and therefore a reason why the stars, which are contained in its space, should be finite in number."

Compared to the fiery enthusiasm of Bruno in preaching the infinity of the universe and the absence of any center (a center which, let us stress it, persists in the Copernican system), Newton's argument appears somewhat utilitarian. Anyway, finite or infinite, the Newtonian (static) universe is unstable. The difficulty could not be overcome and Newton abandoned cosmology. To solve this difficulty, or better to say, to free oneself from it, it will be necessary to wait for the beginning of the twentieth century and Einstein's two theories of relativity, the modern physical theories of space and time that have replaced the absolute space and time of Newton.

7 Albert Einstein, the New Magellan

By the end of the 19th century the aether had become topical in physics, not as the fifth element of the superlunary world, but as the medium where electromagnetic waves propagate. Ironically, ether also reintroduced a strange asymmetry in the physical world, similar to that of Ptolemaic cosmology. There were on the one hand the laws of mechanics, valid in all inertial frames, and, on the other hand, the laws of Maxwell's electromagnetism, valid only in the referential of the aether. But all attempts to detect the aether experimentally, including the experiments of Michelson and Morley, had failed.

In 1905 Albert Einstein, a young third-class employee at the *Federal Intellectual Property Office* in Bern, cut short all efforts to give an explanation to these negative results in the framework of Newtonian physics: aether does not exist, he said, and *the laws of physics are identical in all inertial reference systems.* This is the statement of

[7] But also to the infinite power of God that produces all the infinite effect that it can produce.

the (meta-)principle of relativity of Einstein. It is a law on the laws of physics. And since the speed of light appears in Maxwell's equations as a constant of physics, its invariance follows: it does not depend on the reference frame (that is, the speed) of the experimentalist who measures it.

The speed of light is therefore a conversion factor: time can be measured in metres and distances in seconds (or light-years): "Accordingly we can express the essence of this postulate very tersely in the mystical formula: 300,000 km = 1 s." Space and time, thus, are the same thing (up to a minus sign): "Henceforth, space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality" (H. Minkowski, 1908).

Without having taken this fundamental step, that is to say without having merged space and time into the Minkowski spacetime [17], the question of a scientific cosmology could not be asked. But the crucial ingredient was still missing: gravity. Indeed, Newton's law of universal gravitation does not find its place within Einstein's relativity of 1905 because it presupposes an instantaneous action at a distance—this is also very exactly what shocked Newton. What to do next?

In 1907, Einstein was still employed at the *Patent Office* in Bern (his work as "shoemaker"as he ironically calls it). In his spare time, he reflects upon the way to integrate gravitation into the new relativistic framework. Legend has[8] it, that one morning a Bernese newspaper reported the story of a worker who fell from the roof of a building under construction. Suddenly, Einstein had *den glücklichsten Gedanken*, the happiest idea of his life: "I was sitting on a chair at the Federal Office in Bern when suddenly an idea came to my mind: a person in free fall does not feel his own weight. I was amazed. This simple thought made a profound impression on me. It pushed me toward a new theory of gravitation."

A man who falls freely does not feel his weight and sees his tools floating around him as if there were no gravity: this is Galileo's Principle of Equivalence. And, on the contrary, a man in a spaceship that accelerates into empty space would feel heavy on his seat; if he dropped an object, it would fall to the ground on a parabolic trajectory. There would be no way to distinguish the effects of gravity and acceleration locally: this is Einstein's Principle of Equivalence. Einstein's principle continues and accomplishes the discourse initiated by Galileo and Newton three centuries before.

An immediate consequence of Einstein's principle of equivalence is that gravity must bend the rays of light in the same way as it curves the trajectories of material bodies that would otherwise be rectilinear. A new conception of gravity then emerges: gravity is not to be thought as a force that acts at a distance but rather as a geometric property of the spacetime; the gravitational attraction is a manifestation of the spacetime curvature. The spacetime around the sun is curved and the planets follow geodesics in a non-Euclidean geometry. The resulting orbits are ellipses, with the notable exception of Mercury, whose trajectory is not closed. This is Einstein's answer to the question *a quo moventur planetae.*

[8] Curiously, the three great moments of the history of gravitation are all accompanied by a mythical narrative (all three presumably imaginary): Galileo who drops objects from the top of the leaning tower of Pisa, Newton and the apple falling on his head, Einstein and the worker falling off the roof.

And which is the agent curving the spacetime? The matter contained in the universe, or, more precisely, its energy-momentum content. Einstein's equations of 1915 tell exactly how this happens. They mathematically translate the following idea into a set of equations:

Spacetime Curvature = Energy-Momentum of Matter.

After writing his equations Einstein naturally turned to cosmology and tried to apply them to the entire universe. This is the founding act of a new science, the modern scientific cosmology. The new, very bold idea is that a cosmological model, a model for the universe, corresponds to a global exact solution of the equations for the geometry of spacetime. Einstein thus resumed the discussion initiated by Newton in the *Principia* on the nature of space and inertia.

As very often, Einstein's preoccupation in writing his *Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie* is at first epistemological: at the very beginning of the paper, he declares that [18] "In a consistent theory of relativity there can be no inertia relatively to "space," but only an inertia of masses relatively to one another. …If I have a mass at a sufficient distance from all the other masses in the universe, its inertia must fall to zero." Einstein will later call this property the *Mach principle*.

The metric structure of the universe also must be entirely determined by matter; no matter, no universe and therefore no absolute space. But general relativity still keeps a remnant of absolute space by the boundary conditions that must be specified at infinity to determine the geometry of spacetime. These conditions are a clear violation of the Mach principle. Here is Einstein's "crazy idea"[9] to solve this problem: a spherical universe.

A spherical universe? Again?

Einstein's spherical space is very different from that of Aristotle and Copernicus which is a spherical three-dimensional bubble bordered by the two-dimensional spherical surface of the fixed stars. Instead, one must imagine the sphere of Einstein as a "hypersurface": a three-dimensional sphere immersed in a four-dimensional Euclidean space (for lack of imagination, we can think of the leather surface of a football sphere, but having three dimensions rather than two). It is therefore a non-Euclidean three-dimensional spherical geometry, that is to say a curved geometry of the space. This sphere obviously has no center, or rather it has its center every-

[9] *I have completely abandoned my views, rightfully contested by you, on the degeneration of the metric. I am curious to hear what you will have to say about the somewhat crazy idea I am considering now.* A. Einstein. Letter to W. De Sitter of February 2, 1917.

I have again perpetrated something relating to the theory of gravitation that might endanger me of being committed to a madhouse. A. Einstein. Letter to P. Ehrenfest, February 4, 1917.

where[10] and any point is equivalent to any other point. It has no boundary either; and therefore: no boundary, no conditions on the boundary.

Einstein contradicts Bruno's principle of sufficient reason and finds a reason (unsuspectable for Bruno, although somewhat utilitarian) for the finitude of the universe. But he agrees with Bruno about the absence of a center of the universe and also about the absence of an edge, which has now become possible even for a finite universe.

There was also a second guiding principle in Einstein's cosmological research, a principle that was very reasonable in 1917: the universe is static and therefore the geometry of the universe must not change as time goes by. Einstein then faced an unforeseen difficulty: his *General Theory of Relativity* of 1915 does not allow for static solutions with a spherical spatial geometry. At this point he got the idea that would go down in history as his *biggest blunder*—the biggest mistake of his life: to add to his equations a constant term, the cosmological constant, designated by the letter Λ of the Greek alphabet:

$$\text{Spacetime Curvature} + \Lambda = \text{Energy-Momentum of Matter.}$$

The cosmological constant acts as repulsive gravity and can counteract the gravitational attraction. Adding this term is nevertheless—Einstein says [18]—"an extension of the equations which is not justified by our real knowledge of gravitation." With this additional term Einstein's equations admit a perfectly Machian spherical and static solution: it is the static model of Einstein of 1917. In this model, the radius of spherical space is directly proportional to the total mass of the universe. And so, if there is no mass, there is no universe either!

It should be noted, however, that a positive curvature of space is possible even if the additional term is not present: "This term is necessary only for the purpose of making possible a quasi-static distribution of matter as required by the low speed of stars." [18]. This commentary indicates that already in his 1917 paper Einstein was aware of the fact that his original equations of 1915 implied a dynamic universe, but he had set aside this possibility.

Today Einstein is sometimes accused of lack of confidence in his original equations of 1915 which made him miss the discovery of the expansion of the universe, foreseen on theoretical grounds by Alexander Friedmann and Georges Lemaître in the 1920s and observed by Edwin Hubble in 1929. But in 1917 the visible universe still coincided with the Milky Way, the nebulae enigma had not yet been solved and the hypothesis of a static universe was perfectly reasonable.

Like the great Magellan,[11] Einstein had left for his cosmological adventure on a false trail based on wrong maps but, like Magellan, he did find the *paso*, he went

[10] *Sphaera infinita cuius centrum est ubique, circumferentia tamen nullibi* : this is the second definition of God that can be read in the *Liber XXIV philosophorum*, an anonymous medieval treatise attributed to Hermes Trismegistus. Nicolas de Cues applies this definition to the universe: The world machine has, so to speak, its center everywhere and its circumference nowhere (*De docta ignorantia*, 1440). Bruno at many different places later takes up the definition.

[11] *Magellan was deceived by his forerunners' mistake when, upon the warrant of Behaim's chart, Schoner's globe, and the unnamed pilots' story, he formed his great design of circumnavigating*

through the dangers of the *estrecho* and opened the way for navigating in an ocean never seen before by the human eyes. Today, a century after the reading the *Kosmologische Betrachtungen* at the *Prussian Academy* on February 8, 1917, our way of thinking at the universe has not changed. Alternative theories of gravitation are multiplying, cosmological models abound, quantum mechanics is added to and plays the cosmological game, but the paradigm set by Einstein a century ago remains intact.

8 The Fate of the Universe

Our tale could stop here. It is however impossible to conclude this story without alluding to the expanding universe and the destiny of the cosmological constant. Einstein had refined his cosmological ideas through a very intense exchange of letters with the Dutch astronomer Willem de Sitter. Shortly after the publication of Einstein's paper, de Sitter published a second solution of Einstein's cosmological equations: a universe without matter made only with the cosmological constant.

De Sitter's solution, which is perfectly anti-Machian, displeased Einstein a great deal, but his attempts to demonstrate that there was a fault somewhere in de Sitter's calculations and that his universe was not empty, were vain. Finally, Einstein surrendered to the fact that the de Sitter universe was indeed an empty (i.e. without matter) regular solution of his cosmological equations but, he said, it was nevertheless a model without physical interest because of its being not globally static.

In fact, until the early 1930s almost no one had taken seriously the fundamental articles of Friedmann (1922, 1924) who made use of the original equations of general relativity of 1915 to describe expanding universes. Einstein was one of the few to have read them, and he even wanted to publish a note about an error in Friedmann's work. But there were no mistakes. In the retraction of his commentary one could have read "…Friedmann's paper while mathematically correct is of no physical significance…" but, fortunately, Einstein deleted this sentence on the proofs of his paper at the last moment before publication.

Lemaître's independent work of 1927 is based on the cosmological equations of 1917 and also describes an expanding universe. Lemaître's understanding was much more deeply physical than Friedman's; in his original paper (in French) Lemaître gives indeed the first description of the Hubble's law, a law which is now rightly called the Hubble-Lemaître's law: galaxies move away from each other with a speed proportional to their distance because of the expansion of the universe. Einstein listened to this explanation at the 1927 *Solvay Congress* in Brussels but again he was not happy. Lemaître claimed that the universe had a history and Einstein, who was a follower of Spinoza, could not believe it. Einstein's remark at Lemaître on

---

the world. The enigma of Magellan is solved as soon as we recognize that he planned and acted in honest error. Let us not underrate the importance of error. Through the promptings of genius, guided by luck, the most preposterous error may lead to the most fruitful of truths. Excerpt from [19].

this occasion remained famous: "Vos calculs sont corrects, mais votre physique est abominable."

Based on his cosmological model Lemaître also conceived the physical idea of the Big Bang, a name invented by Fred Hoyle during a *BBC* program to mock the cosmology of Lemaître's primitive atom—the explosive beginning of space and time. In an article published by the journal Nature the English physicist Arthur Eddington commented this idea: "Philosophically, the notion of a beginning of the present order of nature is to me repugnant."

We are witnessing here a complete reversal of the perspectives of ancient cosmology; any suspicion of an undue and indigestible mixing between physics and philosophy or, worse, theology, must be pursued with maximal force.

On a trip to Pasadena Einstein learned of Hubble's latest observations and was persuaded of the advantages of dynamic models to describe the universe. In two articles published shortly afterwards, Einstein asserts that the original reasons for introducing the cosmological constant no longer existed. Farewell to the cosmological constant.

Einstein's last message to Lemaître in 1947 was: "The introduction of such a constant implies a considerable renunciation of the logical simplicity of the theory... Since I introduced this term, I had always a bad conscience…I am unable to believe that such an ugly thing should be realized in nature." And here is Lemaître's prophetic answer of 1949: "The history of science provides many instances of discoveries which have been made for reasons which are no longer considered satisfactory. It may be that the discovery of the cosmological constant is such a case."

In fact, Einstein himself had been prophetic in 1917 when, in a letter to de Sitter, he wrote that "In any case, one thing is clear. The theory of general relativity allows adding the term Λ in the equations. One day, our real knowledge of the composition of the sky of fixed stars, the apparent motions of the fixed stars and the position of spectral lines as a function of distance, will probably be sufficient to decide empirically whether or not Λ is equal to zero. Conviction is a good motive, but a bad judge."

In 1997, exactly seventy years after its discovery, we indirectly observed the cosmological constant, or maybe something similar that we now call "dark energy". Nowadays everyone believes in its existence and the de Sitter's empty spacetime has become the paradigm of our universe and also its everyday improving approximation.

It is thought that the cosmological constant constitutes seventy percent of the energy content of the universe and that its proportion is destined to increase, if one believes in the standard cosmological ΛCDM (cold dark matter) model. In the end, only the cosmological constant will remain.

La nuit éternelle commence, et elle va être terrible. Que va-t-il arriver quand les hommes s'apercevront qu'il n'y a plus de soleil?[12]

We do not know the answer to Nerval's anguished cry. Perhaps it is better to leave the last word to poetry, as in the beginning of our journey.

[12] Gérard Nerval - Aurélia ou le Rêve et la Vie (1855).

Fernando Pessoa—who was an enthusiastic reader of Giacomo Leopardi—echoes Leopardi's Asian wandering shepherd in his heteronym Alberto Caeiro's *Guardador de Rebanhos* with the following words[13]:

Num dia excessivamente nítido,
Dia em que dava a vontade de ter trabalhado muito
Para nele não trabalhar nada,
Entrevi, como uma estrada por entre as árvores,
O que talvez seja o Grande Segredo,
Aquele Grande Mistério de que os poetas falsos falam.

Vi que não há Natureza,
Que Natureza não existe,
Que há montes, vales, planícies,
Que há árvores, flores, ervas,
Que há rios e pedras,
Mas que não há um todo a que isso pertença,
Que um conjunto real e verdadeiro
É uma doença das nossas ideias.

A Natureza é partes sem um todo.
Isto e talvez o tal mistério de que falam.

Foi isto o que sem pensar nem parar,
Acertei que devia ser a verdade
Que todos andam a achar e que não acham,
E que só eu, porque a não fui achar, achei.

[13] On an excessively clear day, A day when I wanted to work hard not to work on it at all, I saw like a road through the trees, It might have been be the Great Secret, That Great Mystery of which false poets speak. I saw that there is no Nature, That Nature does not exist, That there are hills, valleys, plains,That there are trees, flowers, herbs,That there are rivers and stones. But that there is no whole to which they belong, That a real and true whole Is a disease of our ideas. Nature is parts without a whole.
This maybe the mystery they are speaking of. This is what without thinking or stopping, I understood should be the truth That everyone is trying to find and does not find, And only I, because I didn't try, succeeded.

References

1. C. Linnaeus, *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species cum characteribus, differentiis, synonymis, locis*, Laurentii Salvii, Stockholm (1735)
2. U. Moschella, Brève Histoire de la pensée cosmologique, *Qu'est que c'est la gravité?*, ed. by E. Klein, P. Brax, P. Vanhove. Dunod, Paris (2019)
3. R. Brague, *La Sagesse du monde* (Histoire de l'expérience humaine de l'univers. Paris, Fayard, 1999)
4. Aetius, *Placita philosophorum.* I sec. a C.
5. D. Bradshaw, The logoi of beings in greek patristic thought, in *Toward an Ecology of Transfiguration.* Fordham University Press (2013)
6. Aristotle, *De caelo*, I, 9
7. G. Bruno, *La cena delle ceneri.* Dialogo primo (1584), in *Opere italiane*. Introduzione di Nuccio Ordine. Torino: UTET (2002–2013)
8. N. Copernicus, *De revolutionibus orbium coelestium*, Dedicatory letter to Pope Paul III. Nurnberg (1543)
9. G. J. Rheticus, *De libris revolutionum Copernici narratio prima*. Basel (1541)
10. A. Osiander, in *De revolutionibus orbium coelestium*, ed. by N. Copernicus, Anonymous preface. Nurnberg (1543)
11. P. Feyerabend, Against method: outline of an anarchistic theory of knowledge. Verso Books (1975) – First edition, in *Analyses of Theories and Methods of Physics and Psychology*, ed. by M. Radner, S. Winokur. University of Minnesota Press, Minneapolis (1970)
12. B. Brecht, *Leben des Galilei.* Suhrkamp Bd. 1 (1963)
13. G. Galilei, *Il Saggiatore, nel quale con bilancia esquisita e giusta si ponderano le cose contenute nella Libra astronomica e filosofica di Lotario Sarsi Sigensano, scritto in forma di lettera all'ill.mo et rever.mo mons.re d. Virginio Cesarini acc.o linceo m.o di camera di N.S.*. In Roma, appresso Giacomo Mascardi (1623)
14. O. Rey, *Réparer l'eau* (Stock, Paris, 2021)
15. I. Newton, *Philosophiae naturalis principia mathematica*. Londini, Jussu Societatis Regiae ac Typis Josephi Streater (1687)
16. G. Bruno, *De l'Infinito Universo e Mondi* (1584), in *Opere italiane*. Introduzione di Nuccio Ordine, Torino, UTET (2002–2013)
17. H. Minkowski, *Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern.* Nachrichten der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse: 53-111 (1908)
18. A. Einstein, *Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie*. Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), pp. 142–152 (1917)
19. S. Zweig, *Magellan* (Reichner, Der Mann und seine Tat. Wien, 1938)

The Role of Quantum Mechanics in Modern Cosmology

Gabriele Veneziano

1 Introduction and Summary

A hundred years ago two radical paradigm changes in our understanding of the physical world had already been thoroughly established:

- **Quantum Mechanics (QM)** as the correct description of the atomic world and of the quanta of light, the photons, with which atoms interact. QM marked the end of classical determinism, as emblematically expressed by Heisenberg's uncertainty principle and by Planck's constant $\hbar$ quantifying that minimal unescapable uncertainty.
- **General Relativity (GR)** as Einstein's extension of Newtonian gravity to velocities comparable to the speed of light c. Its advent marked the end not only of absolute space and time[1] but also of an absolute geometry of space-time. It's successful tests (like the deflection of light by the sun or the precession of mercury's perihelion) had just been carried out dissipating any initial skepticism.

For many subsequent decades both QM and GR were developed and extended, scoring success after success. But the two disciplines hardly interacted with each other. They were considered to be applicable to two very distinct physical situations: the extremely microscopic and the extremely macroscopic worlds, respectively.

As atomic experiments further developed into subatomic ones, involving electrons, protons, neutrons and other elementary particles, the necessity of combining QM with Special Relativity (SR) became a necessity around the end World War II. This process went on for about 30 years, culminating in the formulation, in the early

[1] This was already the case for his previous theory, Special Relativity, which is unrelated to gravity.

G. Veneziano (✉)
Theory Department, CERN, CH-1211 Geneva 23, Meyrin, Switzerland
e-mail: gabriele.veneziano@cern.ch

Collége de France, 11 place M. Berthelot, 75005 Paris, France

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_2

seventies, of the so-called Standard Model (SM) of elementary particles and of their mutual, non-gravitational interactions. The generic name for quantum-relativistic theories, such as the SM, is Quantum Field Theory (QFT). The limitation of QFT to non-gravitational interactions was felt unimportant for a long time, given the weakness of the gravitational force among elementary particles.

In parallel, GR pursued its own adventure with striking results, like predicting the existence of gravitational waves and black holes, both beautifully confirmed recently by the direct detection of gravitational waves originating from the coalescence of two black holes to form a third one [1]. Black holes are very massive compact objects, characterized by their mass and spin, and deforming the geometry of space to such an extent that nothing, not even light, can escape from a surface (the "horizon") surrounding them.

The other class of interesting GR solutions concerned the description of the Universe as a whole, as well as its evolution, under the simplifying assumption of an approximate isotropy and homogeneity (i.e. of being roughly the same in every region and direction). It culminated in the so-called hot big bang model of cosmology to be shortly described below.[2] In all these GR developments QM was happily ignored since quantum phenomena were insignificant for the physics of the macroscopic objects of interest to GR at the time.

This parallel development of QM and GR went on undisturbed till the late seventies, early eighties. But something changed, in this respect, during the last forty years or so: this is the topic to be discussed in this contribution.

2 Hot Big Bang Cosmology and Its Shortcomings

In order to understand why around the end of the seventies physicists working in GR and in QFT started to get closer and work together we have to recall some properties of Hot-Big-Bang cosmology.

Under the simplifying, but experimentally well supported, assumption of an approximately homogeneous and isotropic Universe, Einstein's equation can be solved once its matter content is given. An unexpected and startling feature of the solution is that, in general, a static Universe (one that is also the same, on average, at all times) is *not* allowed. The Universe must either expand or contract (or perhaps expand along some directions and contract along others). Physically, this result can be understood to be a consequence of the universal, attractive nature of gravity. An initial static Universe, left to itself, will tend to collapse. In the early twenties this looked like a serious blow for GR since people thought the Universe to be static.

Einstein was upset by this failure and ingeniously found a (we would say today "natural") solution by adding a new term to his equations, the so-called cosmological constant, Λ. Choosing properly its sign it produces a repulsive force that compensates

[2] For some history and more details see Ugo Moschella's and Kai Schmitz's nice contributions to this volume.

the gravitational pull. When, in 1927, Hubble discovered, via the famous red-shift, that the Universe is actually expanding, Einstein retracted his proposal calling it his *biggest blunder* and went back to his original equations.[3]

Once we accept the expanding-Universe solution of Einstein's equations we can retrace its implications for our past history. Here we encounter another surprise: in our past the Universe was obviously smaller than today, and therefore denser. Furthermore, since Einstein's equations relate the density of matter to the curvature of spacetime (meaning a deviation from our usual Euclidean geometry) the Universe was also more and more curved. The problem is that there is no mechanism, within GR, to stop and limit that growth of density, curvature (and temperature). Instead, one finds that all these quantities reach simultaneously an infinite (hence unphysical) value at a finite time in our past, the Big-Bang, an event that would have occurred some 13.8 billion years ago. In jargon one calls such an event a "singularity". Finally, since the solution does not make sense before that instant, one is forced to accept that time did have a beginning, precisely at the Big Bang. This has been the standard cosmological dogma till the early eighties. And, I am afraid, it is the picture that (most) scientists are conveying to the general public even today.

Hot big bang cosmology had some indisputable successes. First of all it predicted, before its accidental discovery by Penzias and Wilson in 1965, the existence of a thermal bath of electromagnetic radiation, the so-called cosmic microwave background (CMB), throughout the Universe. A hot early Universe also explains how light elements, like helium and lithium, were synthesized out of hydrogen and predicts successfully their present abundance. It is therefore undeniable that the Universe, long ago, has been extremely (though not necessarily infinitely) hot. We can also locate the time of the big bang quite accurately: whether it was infinitely—or simply very—hot will not change appreciably the number I mentioned since all those early processes occurred in a tiny fraction of a second! As we shall see, what is not necessarily justified is to call that number the *age of the Universe*.

Actually, here is a first signal that QM can play a role in cosmology. As pointed out first by Max Planck at the beginning of last century any quantum theory of gravity will be characterized by a length or time scale given by the appropriate combinations of c, $\hbar$ and Newton's constant G. These are called Planck length and Planck time and their values in meters and seconds are:

$$l_P = \sqrt{\frac{\hbar G}{c^3}} \sim 1.6210^{-35}\mathrm{m}\ ;\ t_P = \sqrt{\frac{\hbar G}{c^5}} \sim 5.3910^{-44}\mathrm{s}\ ;\ M_P = \sqrt{\frac{\hbar c}{G}} \sim 2.1810^{-5}\mathrm{g}\,, \tag{1}$$

where, for later convenience, we have also introduced the Planck mass in grams.

Precisely because the Universe was so dense and curved at its birth, we can ask whether one can reliably calculate its early evolution while neglecting quantum effects i.e. by using GR as it is. Even in the absence of a satisfactory theory of

[3] In hindsight Einstein's blunder was not so much in introducing Λ, but in having to fix it at precisely the value that leads to a static Universe. Any tiny deviation from that value would either lead to an expansion or to a contraction: in modern terminology Einstein's Λ had to be fine-tuned to an extremely high accuracy.

quantum gravity one can argue that this is only justified from a few Planck times after the Big Bang on. In other words, statements based on GR about the Universe at times of order t_P from the Big-Bang, or earlier, and at fortiori the Big-Bang itself, cannot be taken seriously.

But, we may ask, what's wrong with Hot Big Bang cosmology if we decide to start our history of the Universe a few Planck times after the Big Bang and use GR from then on? This has been the pragmatic attitude of cosmologists till the seventies. Identifying what's wrong with that cosmological scenario led to the development of inflationary cosmology in the eighties.

The shortcomings of the Hot Big Beng scenario can be all ascribed to a general property of ordinary matter: the fact that gravity is an attractive force. That implies that the expansion of a Universe containing ordinary matter tends to decelerate since the gravitational attraction resists the expansion.[4] I will only mention two serious shortcoming of Hot-Big-Bang cosmology, both related to the above-mentioned property of the hot big bang scenario.

- **The flatness problem**
  Present observations of the CMB show that today the geometry of space (on large scales) is very close to being Euclidean. However, it is easy to show that, for a *decelerating expansion*, deviation from (the flat) Euclidean geometry tends to increase with time. Inserting the appropriate numbers one finds that, unless the Universe was already Euclidean to one part in 10^{30} a few Planck times after the big bang, it would be impossible to understand why it is now observed to be Euclidean up to, at most, one part in a hundred.
- **The isotropy/homogeneity problem**
  The CMB has a frequency spectrum typical of a thermal distribution at a temperature of about $2.7K$ (meaning 2.7°C above the absolute zero). As I already mentioned, predicting the existence of the CMB, including a fairly good estimate of its temperature, was one of the successes of HBB cosmology. So what's the problem? The problem is that this temperature, when measured along different directions in the sky, looks exactly the same.[5] For a *decelerating expansion*, however, one finds that the CMB radiation coming from, say, opposite directions in the sky, originated from regions of space that were *all the time* so far separated from each other that there was no way they could have ever interacted, exchanged energy, and thermalized. In other words, in order to explain the isotropy of the CMB, the newly born Universe had to be already extremely isotropic.

Both shortcomings can be best regarded as *extreme fine tuning* problems, meaning that they can only be solved if one chooses initial condition (say a few Planck times after the big bang) corresponding to: i. an extremely small spatial curvature, and ii. an extremely homogeneous/isotropic Universe.

---

[4] Conversely, gravitational attraction would accelerate a contracting phase and this is why one gets an infinite contraction rate at the big bang itself.

[5] The first experiment to reveal that this was only true up to variations of tens of microKelvins (i.e. differences of about one part in a hundred thousand) was the celebrated COBE satellite experiment in 1992 [2].

3 A Simple Solution: Inflation!

It will not have escaped the attentive reader that the source of the above difficulties is closely related to the one confronted by Einstein when he was trying to get a static Universe from his equations: the attractive Nature of gravity! In order to get an *accelerated expansion* one has to find some repulsive contribution like the cosmological constant Λ introduced—and then rejected—by Einstein.

Indeed introducing a large-enough positive Λ one does achieve, instead of a static Universe, one that has an accelerating expansion. This is, incidentally, one of the favorite explanations for the recent acceleration of the expansion as measured e.g. by the Supernovae experiments [3, 4]. Why then not play that same game in the early Universe?

The problem is that a cosmological constant, as its name says, corresponds to an energy density that remains constant in time, in spite of the expansion. Since, instead, other forms of energy density (matter, radiation) decrease with time, an initial largish cosmological constant would have dominated every other form of energy since the early days after the Big Bang. And this is incompatible with what we know about the history of the Universe.

Fortunately however, field theories, even at the classical level, offer better alternatives. A very simple and popular one is to invoke the potential energy stored in a scalar field when the value of the field is not the one corresponding to the minimum of the potential. QFT abounds of such fields, a famous example being the Higgs field. Today the Higgs field has a non-vanishing value corresponding to where its potential energy is as small as possible. Its non-zero value generates masses for most of the elementary particles we know. But one can argue that, when the temperature of the Universe was very high, the Higgs field was actually zero. At the same time its potential energy was larger than it is today. Well, that positive potential energy behaves precisely like a cosmological constant and can produce an accelerating expansion. With an important difference: unlike a cosmological constant put in by hand, the Higgs field can evolve (e.g. as the Universe expands and cools down) and its potential energy can eventually disappear in favor of other more conventional forms of energy (e.g. radiation).

The Higgs field was just an example (yet taken as a serious possibility by some authors) to illustrate how a scalar field can play the role of a "dynamical" Λ, doing its job when it's needed, and then kindly disappearing when it's no longer necessary. This accelerated (quasi exponential) expansion of the Universe has been dubbed inflation, and the hypothetical scalar field responsible for it is called the inflaton.

A long enough inflationary epoch completely solves the problems we mentioned: if there were some initial deviations from Euclidean geometry (called spatial curvature in GR terminology) it is wiped out almost completely so that, even after a long decelerating expansion, it is still small (but perhaps non zero and measurable) today. By the same token, inflation can wipe out any initial inhomogeneity or anisotropy by

stretching space so much that any initial ripples are now far beyond our cosmological horizon. It looks as if all problems have been cured, but that's were QM makes its entry on the scene, as we will now discuss.

4 The Crucial Role of Quantum Mechanics

As we have just explained inflation is very efficient for smoothing out any inhomogeneity already present at its onset. However, without the help of quantum mechanics, inflation, if it lasts long enough to solve the fine-tuning problems of hot BB cosmology, does this too efficiently: it produces an exactly homogeneous Universe.

On the other hand, in order to generate the large-scale structures we see in the sky, one has to start with a small, but finite, amount of inhomogeneities. The level of inhomogeneities we have measured in the CMB is roughly of the right order of magnitude to be able to do the job as the Universe keeps expanding and cooling down. This is again due to the attractive nature of gravity making it easy for over-dense regions to grow by accretion of surrounding matter. But what can then produce those small initial fluctuations in the CMB temperature? The mere assumption of a long inflationary phase cannot.

Quantum mechanics, instead, does just that. Although *initial* fluctuations are very effectively wiped out, quantum mechanics keeps creating small-scale fluctuations *all the time* during inflation. Like any other distance, the wavelengths of these fluctuations are stretched during the inflationary era. They are also amplified as soon as their wavelength exceeds a certain length scale (inversely) related to the energy density during inflation (which is roughly constant). This is one important free parameter: it can be called the scale of inflation l_i.

In Fig. 1 we show, in a cartoon-like style, how the initial classical perturbations (wiggles in green), as well as some quantum fluctuation (wiggles in red) produced in the earlier stages of inflation, are stretched beyond our present horizon and how, instead, quantum perturbations generated at sufficiently late times during inflation (also in red), are still within our visible Universe.

Given a specific model of inflation, the amount of fluctuations generated at different wavelength can be computed. Their magnitude is fixed by quantum mechanics in terms of the ratio of Planck's length (which, as indicated in (1), contains Planck's constant) and the above-mentioned scale of inflation l_i which, instead, has a completely classical meaning. Hence, roughly, in order to get perturbations at the level of one part in 10,000, we need a scale of inflation of about $10^5 l_P \sim 10^{-30}$m., still a very short length scale! In turn this will fix the value of the inflaton's potential energy during inflation (which is also sometimes referred to as the scale of inflation).

Another parameter that has to be fixed is the slope of the inflaton potential: it has to be sufficiently small for inflation to last for a long time and for the scale l_i to increase very slowly during inflation (as indicated in Fig. 1). This produces an almost scale-invariant spectrum with a slight "red tilt" (a slight preference for longer wavelengths over smaller ones). In conclusion, up to adjusting a few parameters,

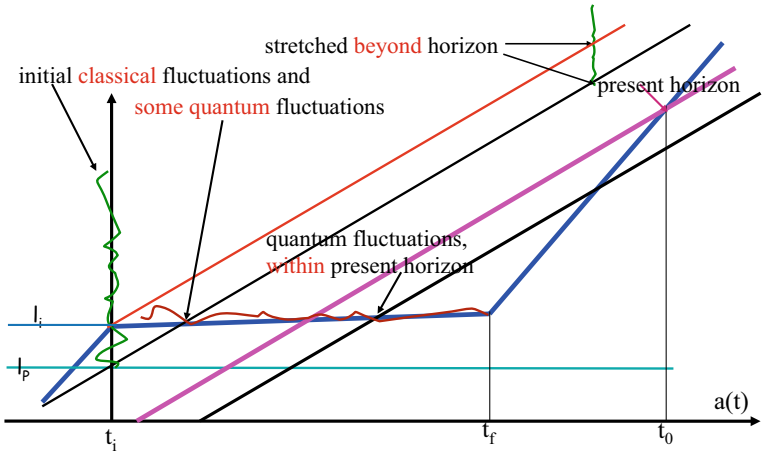

Fig. 1 Kinematics of inflation and of perturbations therein. The horizontal axis represents time in terms of the so-called scale factor $a(t)$, that tells us how physical distances are stretched by the expansion. This is why different wavelengths evolve according to straight, parallel lines. t_i, t_f, t_0 represent the beginning of the inflationary epoch, its end (to be associated with the Big Bang), and the present time, respectively. On the vertical axis we have other relevant length scales: the Planck length l_P and the "inflation scale" l_i. The thick blue line represents the horizon size (how far in space one can communicate) at different times. Its value at t_0 (our present horizon) limits the range of present observations

inflation, together with a crucial help of quantum mechanics, can explain in an amazingly precise way the temperature fluctuations of the CMB as measured very accurately by the PLANCK satellite experiment (see Fig. 2) [5].

But we are not done yet …We need to produce the CMB itself if we want to explain its temperature fluctuations! As we have already discussed the CMB is a left-over of a hot Big Bang. Where is the hot big bang in the inflationary scenario? If it preceded inflation its consequences got wiped out by inflation. Indeed, inflation is an adiabatic expansion that cools the Universe down to essentially zero temperature. The only way out is to reheat (or just heat if it was always cold) it *at the end* of inflation so that it gets gently cooled by the later expansion.

This is the well-known reheating issue in inflationary cosmology. It is solved by dissipating the inflaton's potential energy through some quantum irreversible process, such as quantum particle creation in an external field. I like an analogy with a waterfall in which there is a lot of potential energy stored upstream of the fall, which gets converted into heat (or electricity if we are so smart to use it) as it goes down the fall. Achieving a large enough reheating temperature is one important constraint on models of inflation.

This is the second, equally important intervention of QM in inflationary cosmology. Together they produce, out of a fairly generic initial state, a hot big bang with a sufficiently high (but finite!) temperature and the right amount of fluctuations to seed the large scale structures within our observable Universe!

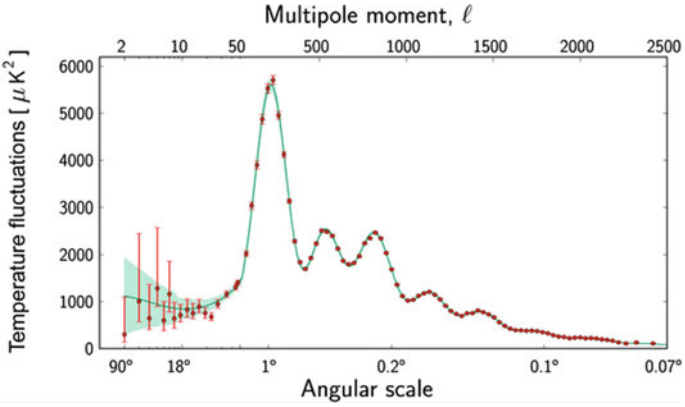

Fig. 2 The PLANCK satellite spectrum of CMB temperature fluctuations as a function of the angular difference between two directions in the sky (shown on the horizontal axis). With a few adjustable parameters, inflation accurately explains the very non trivial structure of these fluctuations including its famous "acoustic peaks". At large angular scales statistical fluctuations prevent any reliable test [5]

This line of reasoning led, around the turn of the millennium, to the formulation of what can be called the Standard Model of Gravitation and Cosmology. It is based on General Relativity, a crucial input from quantum mechanics, and the addition of two *dark sectors*. One dark sector we have already mentioned: it is the invisible (hence dark) energy responsible for the recent acceleration in the expansion of the Universe. The second, instead, is a form of massive matter that interacts with ordinary atomic matter only gravitationally. As explained in great detail in Kai Schmitz's contribution, dark matter is necessary for explaining many features of our Universe, in particular the formation of large scale structures from the seeds we see in the CMB.

The cosmological component of the Standard Model of Gravitation and Cosmology is modestly called "The Concordance Model of Cosmology". A very appealing paradigm indeed, although we still do not know the real nature of its dark sector and we are still far from fully placing it into a grander picture of all known particles and interactions.

5 The True Place of the Big Bang in Modern Cosmology

It clearly emerges from our discussion that the place of the Big Bang in the history of the Universe has moved from its original place and that its very meaning has changed. To put it simply: the Big Bang is no longer the singular beginning of time. It is nothing but the moment (or the process) at which the Universe, after its extreme cooling due to inflation, reheats up as a result of quantum dissipative processes. Thus,

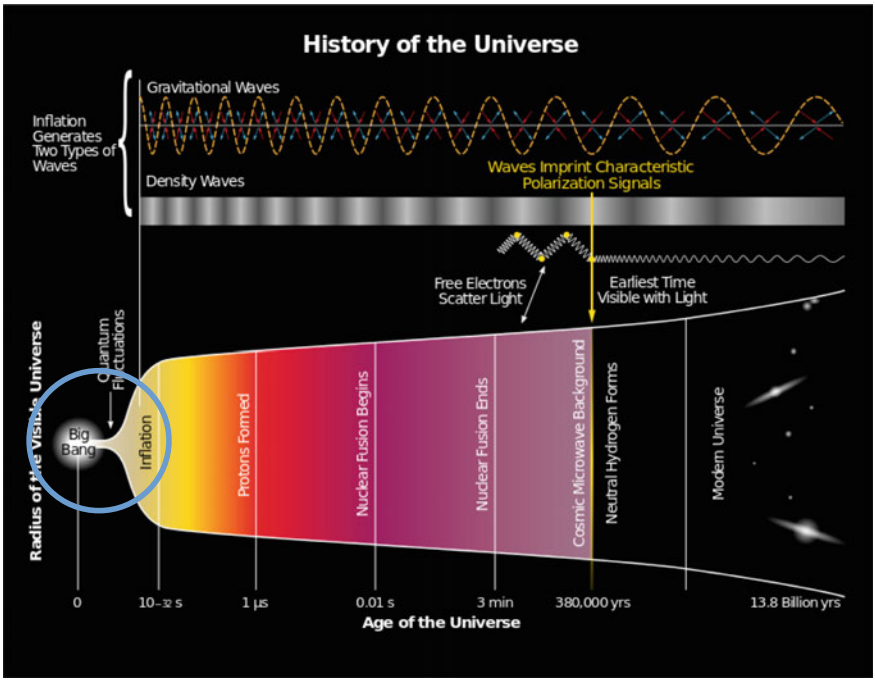

Fig. 3 A popular conventional way of depicting the history of the Universe in which the Big Bang is (mis)placed before inflation (Credit BICEP2 Collaboration)

standard cartoons that illustrate the history of the Universe, such as the one shown in Fig. 3, need to be revised.

Furthermore it is now legitimate, and even physically relevant, to ask: what happened before the Big Bang? And we even know part of the answer, at least for a certain lapse of time before the new Big Bang: there was an inflationary phase and we can study it through the imprint it left (by the quantum fluctuations produced during inflation) on the CMB and on the large-scale structure of the Universe. Studying what happened before the Big Bang has become a physical, *no longer* a metaphysical, question.

An even more direct look at what went on before the Big Bang will hopefully come in the not-too-distant future, from the detection and study of gravitational waves produced during inflation. Gravitational waves, unlike the electromagnetic ones of the CMB, travelled undisturbed even when the Universe was very hot and charged particles were not yet combined to form neutral atoms (this is why it it is impossible to look at the CMB beyond the time of recombination).

The inflationary scenario (or the concordance model of modern cosmology) does not represent, however, a fully self-contained history of the Universe. In particular, they leave open the question of how, when and where inflation started.

We understand pretty well what are the conditions to be satisfied—in a certain region of space and for a certain interval of time—in order for that region to undergo a long inflationary phase and to become, today, as large as our visible Universe. We understand much less the global structure of the Universe. It could have inflationary patches of different kinds with non-inflating regions separating them. Also, each distinct inflationary patch could have different physical properties (the so-called "Multiverse").

In connection with these deeper questions too, QM is bound to play a crucial role: presumably, a full fledged quantum theory of gravity and of the other forces will be needed in order to make progress. So far, an approximate semiclassical approximation was sufficient provided the scale of curvature during inflation was sufficiently small in Planck units. And we know that this was the case during the last phase of inflation because of the smallness of the quantum fluctuations we measure in the CMB. But, if we want to go back even further, a full quantum theory of gravity is very likely needed.

At present, the leading candidate for such a consistent theory is (super)string theory. It is not possible, within the space at my disposal, to even try to describe in any detail what string theory is. It suffices to say that it is a candidate unified quantum theory of all forces and elementary particles based on three basic ingredients (the first two of which we have already discussed and used):

- Quantum mechanics
- Special Relativity
- The postulate that *all* elementary particles are one-dimensional objects, strings. These come in two kinds, open and closed. They are characterized by a single dimensionful parameter, the string tension T (its energy per unit length).

Note that we have *not* included GR in the above assumptions. In the same way that one can define a Planck length, a Planck time, and a Planck mass via Eq. (1), one can define a string length, a string time, and a string mass by:

$$l_s = \sqrt{\frac{c\hbar}{T}} \;\; ; \;\; t_s = \sqrt{\frac{\hbar}{cT}} \;\; ; \;\; M_s = \sqrt{\frac{\hbar T}{c^3}} \tag{2}$$

Notice that these "string quantities" go to the corresponding Planckian ones if we identify the string tension T with $\frac{c^4}{G}$. Is this the way string theory represents gravity? Not quite. In string theory the quantities appearing in (2) are the most basic ones, as we shall see in a moment.

By contrast, Newton's constant G (and thus Planck's time, length and mass) are not as fundamental. They are related to the string quantities through the so-called string coupling g_s a dimensionless number that controls the strength of *all* interactions. It is roughly related to the famous fine structure constant $\alpha \sim 1/137$ of QED. Because of this value of g_s the string length is not expected to be more than a couple of

orders larger than the Planck length[6] but this is sufficient for it to have a huge impact on quantum gravity. At the string mass/energy scale all four forces unify and have a strength given in terms of g_s^2. At low energy string theory is well described by its "massless modes" hopefully to be identifiable with the particles of the standard model or of some extension of it, while at high-energy and short distances it will differ from an ordinary quantum field theory in essential ways.

The existence of these massless modes is of course phenomenologically crucial. It is actually a consequence of QM since, classically, the only massless strings would be point-like. Instead quantum mechanically a string can have a finite size and yet be massless and even carry non zero angular momentum. These latter strings of spin up to 2 would represent the carriers of all known long range interactions (including gravity mediated by the spin-2 graviton this is why GR comes out of string theory!) and possibly more …

The finite size of strings, on the other hand, manifests itself through modifications of the predictions of QFT when scales of order l_s or shorter are probed. This is precisely what happens when the Universe, as described by GR, has an age of the order of t_s. Since $t_s > t_P$ this means that string modifications intervene *before* one enters a Planckian regime.

Several gedanken experiments have been considered in order to study this *Stringy* regime and have confirmed the softening of GR expectations at short distances. They are sometimes described by an effective Generalized Uncertainly Principle (GUP) reading [6]

$$\Delta x \geq \frac{\hbar}{\Delta p} + \frac{c\Delta p}{T} \geq 2l_s\,, \tag{3}$$

in which the first term is the usual Heisenberg's uncertainty, the second its "stringy" extension, and the last inequality is easily proven. Here we see l_s emerging as a minimal measurable length. Armed with these rough notions about strings, let's go back to the question we asked earlier.

6 Before Inflation …

General theorems proven by by S. Hawking and R. Penrose in the seventies state that, within General Relativity, an initial singularity is unavoidable. However, "before" (as we go backward in time) reaching the singularity we necessarily encounter a situation in which certain physical quantities (density, temperature, curvature of spacetime) reach values of $O(1)$ when measured in the string units of (2). Because of the (small but sufficient) hierarchy between string and Planck units, string effects intervene *before* one reaches the regime in which quantum gravity corrections go out of control.

[6] For the learned reader: there is here an amusing analogy with the theory of weak interactions where there is a fundamental mass scale given by the $W - Z$ masses and another, more phenomenological one, associated with Fermi's constant G_F. Also there the ratio of the two scales is given by a coupling whose value is not very far from 1.

String effects, on the other hand, are well known to tame the bad behavior of quantum gravity at short distances (solving e.g. its "non-renormalizability" problem). It is also known, for instance, that string theory sets un upper limit (of course of $O(1)$ in string units) to temperature (the so-called Hagedorn temperature). And even measuring distances (and time intervals) smaller than l_s (or t_s) don't make sense in view of (3).

We may thus reasonably guess that, within string theory, there is no place for a singularity taking place before inflation. Rather, the singularity and what follows it for a time $O(t_s)$ should be replaced by a *stringy phase* during which a conventional classical description of space time is no longer valid. String theory allows for solutions that do not correspond to any smooth classical geometry and yet are perfectly well defined. One such solution could represent the true beginning and evolve at later times into a more conventional inflationary epoch.

There is, however, an interesting alternative to this possibility within string theory, going under the name of Pre Big Bang (PBB) scenario. Born in the early nineties it has been the prototype of a whole class of cosmologies now known as *bouncing* cosmologies. In this class of scenari, the usual cosmological expansion (with or without an inflationary epoch) is preceded by a contracting phase which, through a bounce, turns into an expansion. Such a behavior is forbidden in Einstein's gravity but quantum/string effects could possibly induce such a bounce.

Many scenarios of this sort have been proposed, but I'll limit myself to describe the original one since it is deeply rooted in some novel symmetries characteristic of string theory. These symmetries, known as T-dualities, involve in an essential way a scalar field which is ubiquitous in string theory as an inevitable partner of the gravitational field: it is called the dilaton and plays a very important role in string theory. The above mentioned string coupling g_s is not a God-given number but is itself a field, the dilaton. Even if we know, through its connection with the fine structure constant, that the value of this fields cannot have changed appreciably since many billions years, it is all but excluded that it may have evolved in primordial cosmology and, a fortiori, before the hot Big Bang.

Actually the symmetries of string cosmology associate with the usual decelerated expanding solution one of accelerated expansion (i.e. of inflation), the acceleration being driven by the evolution of the dilaton from the regime of extremely weak coupling to the one corresponding to its present value. This pre-bounce phase has been dubbed dilaton-driven inflation (DDI). While the coupling grows so does the ratio l_P/l_s so that, while the size of the Universe keeps growing all the time in units of l_s, it undergoes a contraction before the bounce if sizes are measured in units of l_P. What remains true, independently of the units adopted, is that the curvature of spacetime grows before the bounce and decreases afterwards: the curvature undergoes a bounce! Details of the bounce itself will only be describable in a full quantum string theory context to be still fully developed.

A very ambitious possibility is that the DDI phase plays the role of ordinary inflation so that, after the bounce, one goes over directly to a standard hot big bang cosmology (explaining the name given to this scenario). Quantum particle production during the pre-bounce phase has been shown to be able to heat-up the initially cold Universe. A detailed scenario has been constructed (see [7] and references therein),

invoking other fields present in string theory, which is compatible with CMB observations and predicts a vanishingly small B-mode in the CMB polarization. It may also be able to generate seeds for the observed cosmic (intergalactic) magnetic fields whose origin is still very mysterious.

On the negative side, the PBB scenario can (but does not and does not automatically) generate a quasi-scale invariant spectrum of perturbations and may also have difficulties in washing away certain kinds of initial anisotropy. Therefore, a more modest possibility would be that the pre-bounce phase is followed by the above-mentioned string phase and, finally, by a long enough conventional inflationary epoch.

Of course all of this is highly speculative. The good side of the story is that it is no longer a tabu to ask experimentally questions like: What happened before the Big Bang? What happened before inflation? One day we may even find out, scientifically, the answer to a philosophical question that goes back to Augustin, Aristotele, and probably much farther in the history of mankind: *Did time have a beginning*?

References

1. B.P. Abbott et al., [LIGO Scientific and Virgo], GW151226: observation of gravitational waves from a 22-solar-mass binary black hole coalescence. Phys. Rev. Lett. **116**(24), 241103 (2016)
2. G.F. Smoot et al., [COBE], Structure in the COBE differential microwave radiometer first year maps. Astrophys. J. Lett. **396**, L1–L5 (1992)
3. A.G. Riess et al., [Supernova Search Team], Observational evidence from supernovae for an accelerating universe and a cosmological constant. Astron. J. **116**, 1009–1038 (1998)
4. S. Perlmutter et al., [Supernova Cosmology Project], Measurements of Ω and Λ from 42 high redshift supernovae. Astrophys. J. **517**, 565–586 (1999)
5. N. Aghanim et al. [Planck], Planck 2018 results. VI. Cosmological parameters. Astron. Astrophys. **641**, A6 (2020). [erratum: Astron. Astrophys. **652**, C4 (2021)]
6. D. Amati, M. Ciafaloni, G. Veneziano, Can space-time be probed below the string size? Phys. Lett. B **216**, 41–47 (1989)
7. M. Gasperini, G. Veneziano, The Pre - Big Bang scenario in string cosmology. Phys. Rep. **373**, 1–212 (2003)

Modern Cosmology, an Amuse-Gueule

Kai Schmitz

1 Evidence for the Big Bang

1.1 Hubble Expansion

Modern cosmology [1–5] witnessed its dawn about a hundred years ago, in the twenties of the last century, shortly after Albert Einstein had laid the necessary theoretical groundwork in the form of his general theory of relativity [6, 7]. New astronomical observations in 1920s revealed that the spiral nebulae scattered across the night sky are in fact distant galaxies, implying that our Universe is vastly larger than previously thought and our home galaxy, the Milky Way, just one among nearly countless many. This realization lends support to the Copernican principle, the idea that we do not observe the Universe from a privileged vantage point. Together with the observation that the Universe appears to look the same no matter where we point our telescopes, the Copernican principle provides the basis for the cosmological principle: On distance scales larger than galaxy superclusters, the Universe is homogeneous and isotropic; its properties are the same everywhere and in all directions.

General relativity in combination with the cosmological principle almost unavoidably predicts that the Universe must be expanding (or contracting, which, however, is not supported by observations). This theoretical prediction was first derived by Alexander Friedmann in 1922 [8], who noticed that all physical distances in a spatially homogeneous and isotropic Universe must grow in proportion to a universal factor $R(t)$, the cosmic scale factor [9], which is independent of location and only depends on the cosmic time coordinate t. The time evolution of this scale factor is described by two equations that Friedmann derived from Einstein's field equations

K. Schmitz (✉)
Theoretical Physics Department, CERN, 1211 Geneva 23, Meyrin, Switzerland
e-mail: kai.schmitz@cern.ch; kai.schmitz@uni-muenster.de

University of Münster, Institute of Theoretical Physics, Wilhelm-Klemm-Straße 9, 48149 Münster, Germany

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_3

and which are now known as the Friedmann equations. Five years later, Georges Lemaître, not aware of Friedmann's work, independently argued that the Universe might be expanding as well as that an expanding Universe must have a finite age [10]: If the physical distance between two points in space is constantly growing, the two points must have been closer together in the past, until, after rewinding the cosmic clock by a sufficient amount, their separation approaches zero. Lemaître referred to his theory as the hypothesis of the primeval atom, which would later develop into the idea of the big bang (a term coined by Fred Hoyle in 1949 in a BBC radio interview).

According to the big-bang model, the Universe originated from an extremely hot and dense initial state—the further back in time, the hotter and denser—followed by a stage of decelerating expansion causing it to cool down and become more and more dilute. The extrapolation back in time that underlies the big-bang model ultimately breaks down at $t = 0$, where the scale factor goes to zero, $R \to 0$, and the density and temperature of the Universe become infinitely large. The term *big bang* is sometimes used to denote this singularity, even though it merely signals that we have pushed our theory beyond its range of validity. After all, big-bang cosmology is based on general relativity, a classical theory of gravitation, and can therefore be at most trusted up to the Planck era at $t \sim 10^{-43}$ s. At earlier times, quantum-gravity effects are expected to become relevant and regulate the big-bang singularity, albeit in absence of a fully developed theory of quantum gravity, only highly speculative statements are possible at present. In any case, it is clear that the scale factor never vanishes during the stages of the cosmic expansion history that can be described by the standard big-bang model. An infinitely large Universe is therefore always infinitely large in big-bang cosmology. The big bang had in particular no center or spatial origin but occurred everywhere in space simultaneously.

In addition to his theoretical work on the earliest moments of the Universe, Lemaître also pointed out how the expansion of the Universe would manifest itself in astronomical observations. He realized that comoving observers in an expanding Universe would mutually recede from each other, just as raisins in a rising yeast dough recede from each other, even though they do not move with respect to the dough. In particular, he computed that the expected recessional velocities should be linearly proportional to the physical distances between the observers. In the following years, Edwin Hubble was able to observationally confirm this expectation [11], in collaboration with Milton Humason [12] and building on earlier work by Vesto Slipher [13] and Knut Lundmark [14]. To this end, Hubble determined the distances D to a handful of galaxies in our cosmic neighbourhood as well as their recessional velocities v (by measuring the redshift of the light received from these galaxies), which allowed him to identify a universal linear relation of the form $v \approx H_0 D$. This relation agrees with Lemaître's prediction, provides key evidence for the expansion of the Universe and the big bang, and is now known as the Hubble–Lemaître law. The constant of proportionality in the Hubble–Lemaître law is the Hubble constant, $H_0 \approx 70\,\mathrm{km/s/Mpc}$. It corresponds to the present-day value of the Hubble rate $H = \dot{R}/R$, which is defined in terms of the scale factor R and its time derivative $\dot{R}$ and which was much larger than H_0 in the early Universe. The numerical value of H_0 tell us that for each megaparsec that a galaxy is further away from us, its

recessional velocity increases by roughly 70 km/s. Moreover, one can show that its inverse value is closely related to the age of the Universe of about 13.8 billion years. H_0 has always been and still is one of the most important observables in modern cosmology; its precise value is the subject of ongoing research, as we will discuss at the end of this essay.

1.2 Cosmic Microwave Background

Olbers' paradox confronts us with the puzzle of why the night sky is dark—if the Universe were static as well as large and old enough, we would actually expect it to be as bright as the surface of the sun [15]. Big-bang cosmology resolves this paradox. To see this, we first note that the big-bang model does in fact imply a radiation background that is constantly reaching us from all directions in the sky, a theoretical prediction that was first made by Ralph Alpher, George Gamow, and Robert Herman in the late 1940s [16–18]. This relic radiation can be thought of as the afterglow of the big bang and is an immediate consequence of the dropping temperature in the early Universe. Roughly 380,000 years into its existence, the Universe had sufficiently cooled down, such that electrons and protons could begin to form neutral hydrogen. Following this process called recombination [19], photons ceased to undergo Thomson scattering with free electrons, which allowed them to decouple from the hydrogen gas and henceforth propagate freely through the Universe. Photon decoupling thus gave rise to what is known as the surface of last scattering, a two-dimensional sphere with the solar system at its center that encompasses all spatial locations from which relic photons had a chance to travel to us during the last 13.8 billion years after they had last scattered with an electron. Even though the Universe as a whole is probably infinitely large, this surface defines the size of the *observable* Universe, a horizon beyond which we cannot see because of the finite age of the Universe and the finite speed of light. Of course, the Universe has kept expanding since photon decoupling, which is why this horizon is currently located at a distance of about 45 rather than 13.8 billion light years. Besides, the cosmic expansion is also the key to resolving Olbers' paradox. During their cosmic journey towards us, the relic photons are subject to cosmological redshift, i.e., their wavelength becomes stretched because of the cosmic expansion by around a factor 1100, such that they reach us today in the form of microwave radiation. This cosmic microwave background (CMB) radiation [20], while a fundamental prediction of big-bang cosmology, is not visible to the naked eye, which explains why the night sky is dark.

In 1965, less than two decades after it had been theoretically predicted, Arno Penzias and Robert Wilson accidentally detected the CMB in observational data taken with the Holmdel Horn Antenna in New Jersey [21]. This breakthrough marked another triumph of big-bang cosmology, which was crucial in establishing it as the basis of modern cosmology. As expected for a photon gas of thermal origin, the specific intensity of the CMB agrees perfectly with the spectrum of a blackbody, a blackbody at a temperature of around three kelvin, which is why the CMB is also

referred to as 3 K radiation. In fact, the CMB spectrum recorded by NASA's Cosmic Background Explorer (COBE) satellite mission in the early 1990s represents the best blackbody spectrum ever measured [22]. The CMB therefore provides striking evidence for one of the most important predictions of big-bang cosmology, namely, that the Universe must have been in a hot thermal state in its past.

At first sight, the temperature of the CMB blackbody spectrum appears to be completely isotropic across the sky, in accord with the cosmological principle. Upon closer inspection, however, it turns out that it exhibits minuscule fluctuations around its mean value of around three kelvin. These temperature anisotropies are the consequence of density perturbations in the baryon–photon fluid at the time of recombination, i.e., tiny over- and underdensities that are bound to develop into the large-scale structure of the Universe at later times. Big-bang cosmology does not provide an explanation for the origin of these primordial seeds of structure formation. They are instead attributed to the initial conditions of the hot big bang, which means that their generation must be addressed in theories like inflationary cosmology or one of its alternatives (see Sect. 3.1). The most prominent contribution to the CMB temperature anisotropies has an amplitude of a few millikelvin and the shape of a dipole. The standard view is to ascribe this contribution to our relative motion with respect to the CMB rest frame, even though there is an on-going debate as to whether at least parts of it may not have a different explanation that is not of kinematic origin [23, 24]. If the dipole contribution is subtracted from the CMB temperature map, one is left with anisotropies of the order of a few hundred microkelvin; see Fig. 1.

Another important feature of the CMB is that it is linearly polarized because of Thomson scattering around the time of photon decoupling. Similarly to its temperature T, the polarization of the CMB also exhibits anisotropies across the sky. These polarization anisotropies are typically decomposed into two contributions: a curl-less E mode and a divergence-free B mode. Together, the auto and cross correlation power spectra of these different types of anisotropies (TT, TE, EE, BB) encode a wealth of information about the early Universe and the propagation of the CMB photons towards us. Here, temperature and E-mode polarization anisotropies can be sourced by two different types of primordial perturbations: density fluctuations (i.e., scalar perturbations) and primordial gravitational waves (i.e., tensor perturbations). Large-scale B-mode polarization, however, can only be generated by tensor perturbations (on small scales, gravitational lensing converts E-mode into B-mode polarization). Past CMB experiments, including the COBE [22], WMAP [25], and PLANCK [26] space missions, did unfortunately not succeed in finding evidence for such a B-mode signal. Hope therefore rests on future CMB polarization experiments such as BICEP Array [27], CMB-S4 [28], LiteBIRD [29], PICO [30], and the Simons Observatory [31]. The detection of a large-scale B-mode polarization signal by any of these experiments would amount to the discovery of primordial gravitational waves (GWs), a breakthrough in early-Universe cosmology.

In addition to primordial tensor perturbations, the CMB allows one to measure or constrain a large range of further observables. Another important example is the spectral tilt of the primordial scalar power spectrum, n_s, which is a measure for the scale dependence of the primordial density fluctuations in the baryon–photon fluid.

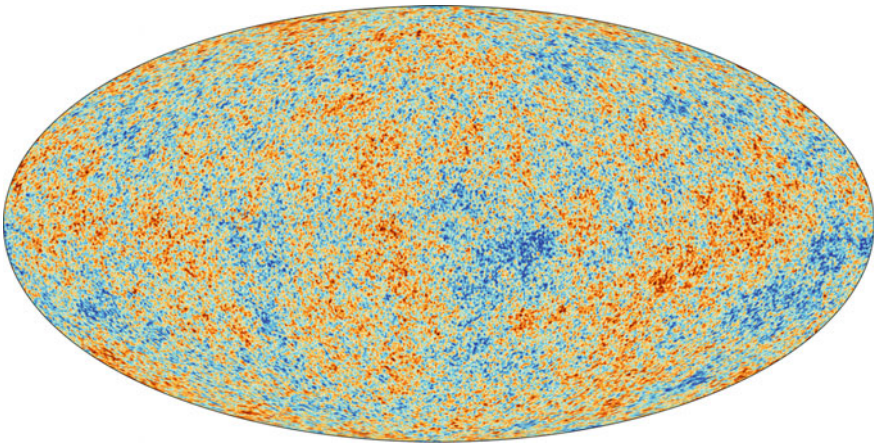

Fig. 1 Baby picture of our Universe roughly 380,000 years after the big bang: temperature anisotropies in the CMB as seen by ESA's PLANCK satellite mission [26]. The mean CMB temperature across the entire sky is $T_{\rm CMB} = 2.72548 \pm 0.00057\,{\rm K}$ [32]; the red spots are hotter than the blue spots by a few hundred microkelvin. Image credit: ESA and the PLANCK Collaboration

Its value is now known to such a precision that a scale-invariant origin of these fluctuations appears extremely unlikely [33]. In other words, whatever physics set the initial conditions of the big bang, it must have involved some scale-dependent dynamical process (such as, e.g., cosmic inflation in the so-called slow-roll approximation, which slightly breaks the scale invariance of perfect de Sitter space). Furthermore, the CMB contains information on the Hubble constant, the spatial curvature of the Universe, its energy budget, and many other quantities such as primordial non-Gaussianities and isocurvature perturbations. A more detailed discussion of these aspects of CMB physics is unfortunately beyond the scope of this essay. Instead, we conclude by emphasizing the importance of the CMB as a key driver of modern cosmology. It is a treasure chest for cosmologists that will continue to play an essential role in the field for many decades to come.

1.3 Primordial Nucleosynthesis

The CMB represents an opaque veil that we cannot see beyond. It is, however, not our earliest probe of the early Universe, which brings us back to the work of George Gamow and his collaborators, whose prediction of the CMB resulted in fact from their interest in a different question, namely, the origin of the chemical elements. Gamow et al. were particularly intrigued by the idea of the big bang because it provides the right conditions for the synthesis of light nuclei from a hot, but gradually cooling plasma of free protons and neutrons. Their seminal work [34] laid the foundation of what is now known as primordial or big-bang nucleosynthesis (BBN) [35, 36] and

which provides us with another window onto the early Universe. Although BBN is a slightly more indirect probe of the big bang compared to the CMB, it allows us to rewind the cosmic clock all the way back to the first second on the cosmic time axis, i.e., a time when the temperature of the Universe was around one megaelectronvolt and hence a million times larger than during photon decoupling.

BBN theory predicts the primordial abundances of isotopes that were formed during the first 20 min or so, notably, deuterium, helium-3, helium-4, and lithium-7. These predictions can be compared to astronomical measurements of the primordial element abundances. Primordial deuterium leaves an imprint in the spectra of high-redshift quasar absorption systems; primordial helium-4 is found in low-metallicity regions of ionized hydrogen; and primordial lithium-7 shows up in the spectra of old metal-poor dwarf stars in the halo of our galaxy. Apart from a slight discrepancy with respect to the lithium abundance [37], which requires further study, one finds very good agreement between theory and observations. This is a remarkable result, especially, in view of the fact that the measured abundances range over nine orders of magnitude. The success of BBN theory thus represents another triumph of big-bang cosmology. Moreover, it leads us to the conclusion that the laws of particle and nuclear physics, which we can test in terrestrial laboratories and which are a crucial input to BBN theory, must have also applied in the early Universe. This is a profound insight into the nature of the physical world. Had we to give up on the universality of the laws of physics, it would become much harder, if not impossible, to make meaningful statements about the cosmology of the early Universe.

Among the most important input for BBN theory are the cross sections for nuclear interaction rates. Progress in the field therefore crucially depends on improved measurements of these cross sections in laboratories, such as, e.g., the recent measurement of the rate of deuterium burning by the LUNA collaboration [38]. Once all nuclear cross sections are set to their measured values, BBN more or less turns into a one-parameter theory, at least as long as one does not introduce any new physics beyond the Standard Model (SM) of elementary particle physics. The only remaining free parameter is then the baryon number density n_b or, equivalently, the baryon-to-photon ratio $\eta_b = n_b/n_\gamma$ at the time of nucleosynthesis. η_b is an important cosmological parameter. It also affects the properties of the baryon–photon fluid at the time of photon decoupling and can hence be deduced from CMB observations. As we find ourselves living in a Universe that does not contain any appreciable amount of antimatter, η_b can also be regarded as a measure of the matter–antimatter asymmetry, or baryon asymmetry, of the Universe, $\eta_b = \left(n_b - n_{\bar{b}}\right)/n_\gamma$, where $n_{\bar{b}} \approx 0$ denotes the present-day number density of antibaryons. BBN theory is able to reproduce the measured abundances of the light elements for $\eta_b \approx 6 \times 10^{-10}$ [39]. Remarkably enough, this value is in excellent agreement with the value inferred from the CMB [33].

Among all of its achievements, this concordance between BBN and the CMB is arguably the biggest success of the big-bang model. It is furthermore an amazing confirmation that, based on the methods of modern cosmology, we are indeed able to reconstruct the history of our Universe all the way back to the first second of its existence. Despite the countless open questions that remain to be answered in

cosmology, it is therefore fair to say that *the big bang did indeed happen*—in the sense that, 13.8 billion years ago, the Universe was indeed filled by a rapidly expanding hot thermal plasma that gradually became cooler and more and more dilute while giving birth to all of the visible matter that permeates the cosmos today. Certainly, the big-bang model cannot be the end of the story. Ultimately, it will need to be embedded into a more fundamental theory that also explains the origin and initial conditions of the big bang, probably in the context of quantum gravity (see Sect. 3.1). However, just like Newton's theory of gravitation will always remain embedded in Einstein's theory of general relativity, the big bang will always remain embedded in the more fundamental theory that is going to succeed it.

The success of BBN theory, next to its important implications for our worldview, also represents a powerful tool for particle cosmology. It enables us, e.g., to constrain the properties of new subatomic particles that may exist beyond the Standard Model (BSM). This includes BSM models that predict the existence of new massless or very light particles, so-called dark radiation [40], which can affect the speed of the Hubble expansion during nucleosynthesis. The amount of dark radiation is typically quantified in terms of a quantity called $N_{\rm eff}$, the effective number of relativistic neutrino species, which counts new radiation degrees of freedom as if they corresponded to new generations of neutrinos. $N_{\rm eff}$ is another important cosmological observable that can be constrained by BBN and CMB observations. With a bit of luck, its precise value may hold the key to the discovery of new particles. Besides that, the outcome of BBN is also sensitive to the presence of heavier particles that decay around or after nucleosynthesis [41]. In some scenarios, such effects may be beneficial, e.g., because they help to ameliorate the lithium problem. In most cases, however, they spoil the success of BBN theory, which then allows one to put bounds on the allowed mass and lifetime of the decaying particle. Regardless of these details, though, it is clear that BBN offers a powerful test that all models of new physics must pass.

2 Concordance Model of Cosmology

2.1 Visible Matter

The excellent agreement between BBN and CMB observations represents an important pillar for the current concordance model of cosmology: the Lambda-cold-dark-matter (ΛCDM) model, which succeeds in explaining the bulk of all cosmological data available today. At its core, the ΛCDM model is a particularly simple model, which, in its minimal form, features only six independent parameters. This simplicity, however, does not come for free; the price to pay is a blatant discrepancy with the Standard Model of particle physics. The crucial point is that the ΛCDM model describes the composition of our Universe in terms of three main ingredients—visible matter, dark matter, and dark energy—all of which call for BSM physics in one way or another. In addition, it relies on initial conditions that cannot be explained

solely based on the Standard Model in combination with Einstein gravity. Cosmology therefore tells us that our understanding of the subatomic world is incomplete. This profound conclusion establishes a deep connection between microscopic and macroscopic physics and serves as a strong motivation for new-physics searches in terrestrial high-energy experiments.

The name of the model, ΛCDM, refers to the fact that it assumes dark matter to be cold and dark energy to be described by a mere cosmological constant Λ; we will discuss both aspects in more detail below. The minimal version of the model moreover assumes a flat spatial geometry, which is only possible if the total energy density of the Universe corresponds to a specific critical value, $\rho_{\rm crit} = 3H_0^2 M_{\rm Pl}^2 \approx 10^{-26}\,{\rm kg/m^3}$ (where $M_{\rm Pl}$ denotes the reduced Planck mass). This critical density allows one to introduce the density parameters $\Omega_b = \rho_b/\rho_{\rm crit}$, $\Omega_{\rm cdm} = \rho_{\rm cdm}/\rho_{\rm crit}$, and $\Omega_\Lambda = \rho_\Lambda/\rho_{\rm crit}$ for baryons, cold dark matter (CDM), and dark energy, respectively. Omitting the negligibly small present-day energy densities of photons and neutrinos, these parameters satisfy by construction $\Omega_b + \Omega_{\rm cdm} + \Omega_\Lambda = 1$ in a flat Universe. Here, the energy densities of baryons and dark matter, $\rho_b \propto \Omega_b H_0^2$ and $\rho_{\rm cdm} \propto \Omega_{\rm cdm} H_0^2$, constitute two of the six free input parameters of the minimal ΛCDM model; the remaining four parameters being: (i) the age of the Universe, t_0; (ii) the spectral index of the primordial scalar power spectrum, n_s; (iii) the amplitude of the primordial scalar power spectrum, A_s; and (iv) the optical depth due to CMB photons undergoing Compton scattering during their cosmic journey towards us, τ.

There is a wealth of observational techniques that allow one to measure or constrain the parameters of the ΛCDM model. In addition to BBN and the CMB, we wish to highlight two of them: the galaxy power spectrum and the supernova Hubble diagram. The former corresponds to the Fourier transform of the galaxy two-point correlation function and can be measured in galaxy surveys. It is an especially useful cosmological probe because it exhibits a characteristic oscillatory feature, so-called baryonic acoustic oscillations (BAOs) [42, 43], which are related to the oscillations in the primordial baryon–photon fluid. The distance scale on which these oscillations occur can be used as a *standard ruler* that determines the size of the sound horizon at the time of photon decoupling. Meanwhile, the supernova Hubble diagram is constructed from measuring the redshift z and distance D of type-Ia supernovae (SNe) [44, 45]. At low redshift, these two quantities are linearly related to each other according to the Hubble–Lemaître law, $z \approx v = H_0 D$. At larger redshift, however, the distance–redshift relation becomes nonlinear, with the exact relation being sensitive to the cosmological parameters.

Together with the CMB temperature and polarization anisotropies, BAO and SNe observations allow for a precise determination of the matter and dark-energy density parameters, $\Omega_m = \Omega_b + \Omega_{\rm cdm}$ and Ω_Λ; see Fig. 2. The result of this measurement is that we live indeed in a flat Universe, $\Omega_m + \Omega_\Lambda \approx 1$, that is dominated by dark energy, $\Omega_\Lambda \approx 0.7$. In the next step, we can combine this result with the measurement of the baryon density parameter from BBN and CMB observations, $\Omega_b \approx 0.05$, which leads us to the fascinating conclusion that the visible matter content of the Universe only accounts for around 5 % of its energy budget. 95% of the Universe is dark, with

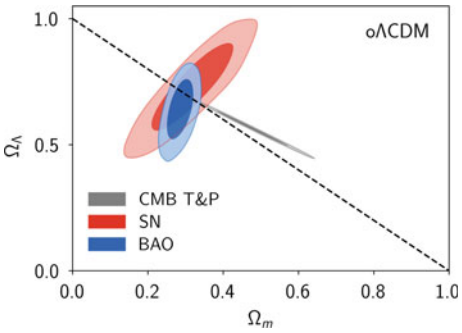

Fig. 2 Constraints on the matter and dark-energy density parameters Ω_m and Ω_Λ based on CMB temperature (T) and polarization (P) data, type-Ia supernovae, and baryonic acoustic oscillations in the galaxy power spectrum [46]. The dashed diagonal line corresponds to a flat Universe, $\Omega_m + \Omega_\Lambda = 1$. The label "oΛCDM" refers to an extension of the minimal six-parameter ΛCDM model that also allows for nonzero curvature (i.e., values of $\Omega_m + \Omega_\Lambda$ larger and smaller than 1). Image credit: The eBOSS Collaboration

around 25% of its energy contained in dark matter and 70% of its energy made up of dark energy.

On the one hand, an energy fraction of 5% in visible baryonic matter may not seem like much, especially when compared to the much larger energy densities of dark matter and dark energy. On the other hand, it turns out to be a remarkably large value when compared to our naive theoretical expectation based on our understanding of the hot big bang. Assuming matter–antimatter-symmetric initial conditions at very early times, we would expect that matter and antimatter almost completely annihilated each other in the early Universe, leaving behind nothing but radiation. In a Universe where such an annihilation catastrophe takes place, the baryon-to-photon ratio would freeze out at a tiny value, suppressed by many orders of magnitude compared to the value that we actually measure in BBN and CMB observations of around $\eta_b \approx 6 \times 10^{-10}$. The fact that we do find a large baryon density compared to this theoretical expectation therefore indicates that the primordial plasma must have exhibited a baryon asymmetry at the time of baryon freeze-out. Roughly speaking, for each billion antiparticles, there must have been a billion and one particle, such that after baryon–antibaryon annihilation the one excess baryon would survive in the plasma. These surviving baryons now constitute the entire visible matter content of the present Universe, while there are no traces of primordial antimatter left.

The baryon asymmetry of the Universe (BAU) is one of the biggest unresolved problems in fundamental physics. Naively, one may be tempted to attribute its origin simply to the initial conditions of the Universe. This supposition, however, would not only be tantamount to a surrender, it would also be in conflict with the paradigm of cosmic inflation (see Sect. 3.1), according to which the hot big bang must have originated from a matter–antimatter-symmetric initial state. It is therefore expected that the baryon asymmetry is the outcome of a dynamical process in the hot thermal phase of the early Universe. This process, commonly referred to as baryogenesis, needs to

satisfy three necessary conditions. As noted by Andrei Sakharov in 1967 [47], any successful baryogenesis scenario must necessarily (i) violate baryon number conservation, (ii) violate charge (C) and charge–parity (CP) invariance, and (iii) involve a departure from thermal equilibrium. These conditions cannot be fulfilled in the Standard Model, which does not feature a sufficient amount of CP violation and does not provide the basis for out-of-equilibrium interactions in the early Universe. The BAU is thus not only a cosmological mystery, it also direct evidence for the existence of new BSM physics. As such, it is the subject of intense on-going research efforts. There is a vast number of baryogenesis models in the literature [48, 49], among which the scenario of leptogenesis [50, 51] is arguably particularly promising. Leptogenesis links the BAU to BSM physics in the neutrino sector and therefore leads to testable predictions for neutrino experiments.

2.2 Dark Matter

Dark matter [52, 53] is roughly five times more abundant than ordinary matter, albeit much less understood. In fact, its very name is not only supposed to indicate that it does not interact with electromagnetic radiation, which is why we are not able to see it with our telescopes. From this perspective, a better name would be *transparent matter* anyway. Instead, the name *dark matter* also refers to our ignorance of its nature. Just like the BAU, dark matter cannot be explained by the Standard Model and therefore provides evidence for new physics.

As a gravitational phenomenon, the existence of dark matter is firmly established—in the sense that, across a large range of distance scales, the visible matter content of the Universe in combination with Einstein gravity does not suffice to explain our observations. We are therefore led to either assume (i) the presence of an invisible form of matter that only makes itself noticeable via its gravitational interaction with ordinary matter or (ii) a modification of general relativity [54]. The second scenario, however, while being successful in many cases, struggles to account for the entirety of existing observations (see Fig. 3 for a famous example), which is why most cosmologists disfavor it. Leaving aside the possibility of modified gravitational dynamics, the data then tells us that dark matter has been a key ingredient of the cosmic energy budget throughout cosmic history.

In the early Universe, dark matter is essential in setting the stage for structure formation. Electrically neutral and nonbaryonic dark matter does not interact with the primordial baryon–photon fluid, which is why it is able to clump together and develop gravitational potential wells already much earlier than ordinary matter. After photon decoupling, the baryonic matter then falls into these pre-existing potential walls, which triggers the formation of the large-scale structure that we observe in the Universe today. This includes highly nonlinear structures such as galaxy halos and clusters, which could not exist if the early collapse of dark-matter overdensities had not provided the appropriate initial conditions. Successful structure formation requires dark matter in particular to be cold, i.e., to have a nonrelativistic velocity at

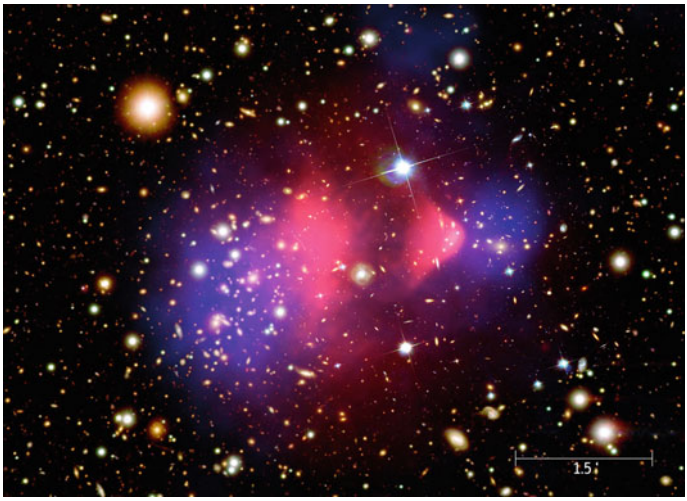

Fig. 3 The Bullet Cluster, a collision of two galaxy clusters that provides compelling evidence for dark matter [62]. Most of the baryonic matter is contained in hot gas visible in X-rays (pink). Gravitational-lensing measurements (blue), however, indicate that most of the system's mass has already passed through the center, offsetting it from the hot gas. This can be explained by assuming that most of the system's mass belongs in fact to (nearly) collisionless nonbaryonic dark matter that does not slow down in consequence of the collision to the same extent as the dissipative gas component. At the same time, the luminous matter (galaxies) traces the gravitational potential induced by dark matter. In theories of modified gravity without dark matter, one would typically expect no offset between the baryonic matter and the mass distribution mapped by gravitational lensing. Image credit: NASA

the onset of the matter-dominated era (see Sects. 3.2 and 3.3). Cold dark matter results in a bottom-up scenario of structure formation, in which matter first collapses into smaller objects, which then gradually aggregate into larger objects. Alternatively, if dark matter were hot, i.e., relativistic at the onset of matter domination, structure would form in a top-down fashion, starting with the formation of large pancake-shaped objects, which then gradually fragment into smaller objects. The three known neutrinos are an example of hot dark matter (as are light BSM neutrinos). However, as cosmological observations and numerical simulations of structure formation strongly support the CDM paradigm, it is clear that hot dark matter can only contribute an insignificant fraction to the total dark-matter density.

In the late Universe, dark matter reveals its existence moreover in a wealth of observations that trace the total amount of mass in gravitationally collapsed structures. This includes in particular the classic observations that historically led to the paradigm shift in favor of dark matter, such as the motion of galaxies in virialized galaxy clusters, which was pioneered by Fritz Zwicky in the 1930s [55, 56], and the rotation curves of spiral galaxies, which were first investigated by Vera Rubin and others in the 1960s and 1970s [57, 58]. Today, these observations are comple-

mented by analyses of gravitational lensing, redshift-space distortions, and many other probes.

It is therefore remarkable that, despite the overwhelming evidence for the existence of dark matter, we still do not know much about its nature at a fundamental level. While we can list some of its key properties (nonbaryonic, cold, dark, etc.) with different levels of certainty, the mass of its fundamental building blocks, e.g., is still unknown. Dark matter may be composed of elementary scalar particles with a mass of the order of 10^{-22} eV, a scenario that goes by the name of ultra-light or fuzzy dark matter [59] and which is motivated by the so-called cusp–core problem of the ΛCDM model. Meanwhile, it is also conceivable that a significant fraction of dark matter is made up of stellar-mass primordial black holes [60]. If some of these primordial black holes have masses a few times the mass of the sun, they might even be partially responsible for the binary black-hole merger events that have been detected by the LIGO and Virgo GW interferometers in recent years [61]. The space of possibilities between these two extremes, fuzzy dark matter and primordial black holes, is filled by a sheer countless number of hypothetical scenarios for the nature and origin of dark matter. In particular, there is no reason why dark matter should consist of only one component. Just like the SM sector, the dark-matter sector may well exhibit a nontrivial structure and encompass several ingredients.

Solving the enigma of dark matter represents a central challenge of 21st-century physics. To achieve this goal, a vast array of experiments and observations is currently in operation or preparation: Numerous laboratory experiments around the world strive to directly detect the exotic subatomic particles that may constitute our galaxy's dark-matter halo; telescopes and satellites look for indirect signatures of dark-matter annihilations or decays in outer space; and particle colliders attempt to directly produce dark-matter particles in high-energy collisions. Other contributions to this book provide more details on the hunt for dark matter.

2.3 Dark Energy

In a Universe filled only with visible and dark matter, one would expect the cosmic expansion to gradually slow down over time (i.e., a negative second time derivative of the scale factor, $\ddot{R} < 0$). Observations of type-Ia supernovae coordinated by Adam Riess, Brian Schmidt, and Saul Perlmutter in the 1990s, however, led to the astonishing conclusion that this type of behavior is not realized in our Universe. By extending the supernova Hubble diagram to large redshifts, the High-Z Supernova Search Team [44] and the Supernova Cosmology Project [45] were instead able to show that the cosmic expansion is currently accelerating (i.e., $\ddot{R} > 0$).

In the framework of general relativity, this observation can only be explained if the stress–energy tensor on the right-hand side of the field equations receives a contribution that acts as a repulsive force. For a perfect fluid with energy density ρ and pressure p, this necessitates an equation of state $p = w\,\rho$ with $w < -1/3$, i.e., a negative pressure with large enough absolute value, $|p| > 1/3\,\rho$. Dark energy is

believed to be such an exotic form of energy with negative pressure, which permeates the cosmos and hence causes its accelerated expansion [63, 64]. In this sense, it represents another gravitational phenomenon that, next to dark matter, cannot be explained in terms of established physics. However, unlike dark matter, which is able to form gravitationally collapsed structures, dark energy is assumed to be largely spatially homogeneous across the Universe.

The concordance model of cosmology attributes the origin of dark energy to the energy density of the vacuum, i.e., empty space. In ΛCDM cosmology, dark energy thus corresponds to an energy density constant in space and time (around 70 % of $\rho_{\text{crit}} \approx 10^{-26}\,\text{kg/m}^3$) that serves as a cosmological constant Λ in Einstein's field equations. This interpretation is motivated by the fact that vacuum energy possesses an equation of state given by $p = -\rho$, i.e., $w = -1$, which leads to accelerated expansion and agrees well with the cosmological data. Indeed, a ΛCDM fit of the combined CMB, BAO, and SNe data fixes $w \approx -1$ within percent-level uncertainties [33].

This phenomenological success is in stark contrast to our microscopic understanding of the cosmological constant Λ. Historically, Einstein had first introduced a Λ term in his field equations in an attempt to construct a cosmological solution that would describe a static Universe. Following the discovery of the expansion of the Universe by Hubble in the 1920s, he discarded this idea, referring to it as the greatest blunder of his life. Thereafter, Λ no longer attracted much interest for many decades. Today, almost a quarter century after the discovery of the accelerated expansion, the situation is again very different and dark energy an active field of research. Nonetheless, no satisfying theoretical explanation is currently in sight. In principle, one may hope that quantum field theory (QFT) would allow one to estimate the size of the vacuum energy density based on the properties of the quantum vacuum. Estimates along these lines, however, turn out to be off by dozens of orders of magnitude, which represents a spectacular and maybe even embarrassing failure of our naive expectations.

The cosmological constant appears in particular to be much *smaller* than what one expects based on simple QFT arguments. Part of the reason for this observation might be of anthropic nature. As famously pointed out by Steven Weinberg [65, 66], Λ values not much larger than the one we observe would be inconsistent with the formation of complex structures in the Universe. The accelerated expansion would simply set in too early during the expansion history of the Universe, leaving no time for the emergence of observers living in complex bound structures. Similarly, the small value of the cosmological constant results in what is known as the cosmological coincidence problem [67]. Apparently, we currently live in a cosmological epoch where Ω_Λ and Ω_m are of the same order of magnitude and the vacuum-dominated stage of expansion has just begun. Is this really just a coincidence, or is there a deeper reason why Ω_Λ must take a value that is neither much smaller nor much larger than Ω_m? In absence of compelling answers to these questions, the origin of the cosmological constant, or dark energy in general, remains one of the biggest puzzles in contemporary fundamental physics.

Alternative explanations of dark energy include, e.g., modified theories of gravity as well as extremely light scalar fields with cosmologically large de Broglie

wavelength. An interesting example of the latter scenario, commonly referred to as quintessence, are models that feature a parity-odd "axion" field coupling to photons. In such axion quintessence models, dark energy can cause a rotation of the CMB polarization angle (i.e., cosmic birefringence). Interestingly enough, first hints for such a signal have recently been reported, although more work is needed in order to better understand the contamination by astrophysical foregrounds [68, 69].

A second property of quintessence is that its equation-of-state parameter can deviate from $w = -1$ and even vary as a function of time. Thus, a clear measurement of $w \neq -1$ would indicate that dark energy is dynamical and rule out the standard scenario of a bare cosmological constant. This could have potentially severe implications for the ultimate fate of the Universe. A cosmological constant characterized by $w = -1$ is expected to lead to an eternally lasting stage of accelerated expansion ($R \to \infty$, $H \to$ const) that eventually results in a big freeze, i.e., a heat death slowly asymptoting towards zero temperature. However, if dark energy should be accounted for by quintessence with a time-dependent equation of state, the future of the expanding Universe would be much more uncertain. In this case, dark energy may at some point even cease to dominate the expansion, which under certain conditions may lead to a big crunch, i.e., a recollapse into a singularity similar to the initial big-bang singularity ($R \to 0$, $H \to \infty$). Finally, one may consider the highly speculative idea of phantom dark energy, i.e., quintessence with negative kinetic energy and $w < -1$. This scenario would result in a steadily increasing energy density of dark energy, which would cause a big rip, i.e., a singularity in the expansion rate ($R \to \infty$, $H \to \infty$) leading to the destruction of all bound objects, even at the subatomic level. All of these ideas are of course very hard, if not impossible to test. Still, future cosmological observations promise to shine more light on the equation of state of dark energy and hence its dynamical origin.

3 History of the Universe

3.1 Initial Conditions

In the previous section, we outlined the main building blocks of the ΛCDM concordance model of cosmology, which now sets the stage for an overview of the expansion history of the Universe during the last 13.8 billion years. To this end, we first make a huge temporal leap from the ultimate fate of the Universe in the distant future back to the first moments of its existence in the distant past and then let cosmic time run forward again.

We begin our cosmic chronology by taking a closer look at the initial conditions of big-bang cosmology, which we already referred to several times in our discussion but otherwise neglected thus far. The crucial point is that these initial conditions, as became increasingly clear during the 1970s, are in fact far from generic. Notably, the cosmological principle, one of the theoretical pillars of modern cosmology, only

applies in big-bang cosmologies with highly fine-tuned initial conditions. Picture, e.g., two CMB photons reaching us from diametrically opposite directions in the sky. Both photons required 13.8 billion years for their cosmic journey towards us and thus stem from points on the surface of last scattering that, because nothing travels faster than a photon, never had a chance to exchange information with each other. Why does the CMB then look nearly the same in all directions? This question is at the core of the so-called *horizon problem*, i.e., the puzzling realization that the early Universe must have been extremely homogeneous across a vast number of causally disconnected patches. Similarly, the CMB and other cosmological probes indicate that our Universe is spatially flat to very good precision. Based on the behavior of spatial curvature dictated by the Friedmann equations, one then concludes that our Universe must have been even flatter, in fact, exponentially close to perfect flatness in the past. This is only possible if the energy density of the Universe is initially chosen to be extremely close to the critical energy density for spatial flatness, $\left|\rho - 3H^2 M_{\rm Pl}^2\right| \lll 1$, which is by no means guaranteed and hence constitutes the *flatness problem*.

A possible solution to these problems emerged around the early 1980s. Alexei Starobinsky [70], Alan Guth [71], Andrei Linde [72], and several other authors noted that a brief period of exponentially fast cosmic expansion could readily yield the required initial conditions for the hot big bang. According to this mechanism, dubbed *cosmic inflation* [73] by Guth, the entire observable Universe was born from a single homogeneous and isotropic domain at early times, a causal patch of microscopic size whose volume was then blown up by a factor of around 10^{90} or even more within a timespan possibly as short as 10^{-36} s. All parts of the observable Universe were therefore once in causal contact, which solves the horizon problem. At the same time, any initial inhomogeneities and anisotropies were stretched out enormously, so that they lie beyond our cosmological horizon today. The same also applies to hypothetical relics from even earlier times, including magnetic monopoles, i.e., point-like topological defects in grand unified theories. If produced during the hot big bang, monopoles would represent the most abundant form of matter in the Universe today, which clearly is not the case. Inflation solves this *monopole problem*, which was actually one of the original motivations for the work of Guth et al., by pushing most, if not all relic monopoles beyond our horizon. Similarly, spatial curvature is diluted during inflation, comparable to the surface curvature of a balloon, which decreases when the balloon is inflated. This solves the flatness problem.

In view of these achievements, we conclude that the evolution of the classical spacetime background during inflation offers an attractive solution to the initial-conditions problems of big-bang cosmology. But that is not all. Remarkably enough, inflation also comes with a built-in mechanism for generating primordial density perturbations at the quantum level. This is because the QFT description of inflation is based on the dynamics of a scalar *inflaton* field (or several such fields). The evolution of the inflaton field background in a shallow scalar potential acts like a time-dependent cosmological constant and hence causes the exponential expansion, while its quantum fluctuations in combination with quantum fluctuations of the spacetime metric give rise to a spectrum of primordial scalar perturbations. These perturbations are continuously produced during inflation from the quantum vacuum, whereupon

they are stretched in length by the expansion, but not damped in amplitude. Inflation thus explains the origin of the temperature fluctuations in the CMB and hence ultimately the origin of the large-scale structure of the Universe. For each given QFT model, one is in particular able to predict two out of the six ΛCDM input parameters based on the microphysics of inflation: the amplitude and the spectral index of the primordial scalar power spectrum, A_s and n_s. The quantum fluctuations of the metric moreover give rise to primordial GWs across a large range of frequencies. This primordial stochastic GW background can leave an imprint in the large-scale B-mode polarization of the CMB and be searched for directly in GW experiments.

We thus arrive at two remarkable conclusions: First, inflation defines a new paradigm, *inflationary cosmology*, which extends the old paradigm of big-bang cosmology by providing a compelling explanation of its initial conditions. A central tenet of this paradigm is that all structure in our Universe is of quantum origin—generated during the first fractions of a second at extremely high energies, when the domain making up the observable Universe today was stretched from a microscopic size, maybe not much larger in diameter than a Planck length, $\ell_{\rm Pl} \approx 10^{-35}$ m, to the size of a macroscopic object, say, an apple. More details on the role of quantum mechanics in inflationary cosmology can be found in the chapter by Gabriele Veneziano. Second, we see that the theory of inflation is capable of making predictions that can be tested in CMB, GW, and other observations.

Next, let us turn to the theoretical description of inflation, which is currently to be considered as a general framework rather than a conclusive theory [74]. The underlying idea of inflation is to use the potential energy density of the inflation field as a form of vacuum energy that can drive a stage of accelerated expansion; however, the microscopic origin of this field remains unclear and leaves room for many speculations. Indeed, many models of inflation, including the early models of old inflation [71], new inflation [75], chaotic inflation [76], etc., have the status of simple QFT toy models. A lot of research during the last 40 years has therefore focused on the possible embedding of inflation into particle physics and string theory, with varying degrees of success.

On the particle physics side, it is, e.g., remarkable that even the SM Higgs field may play the role of the inflation, if it couples in a nonminimal way to the Ricci curvature scalar $\mathcal{R}$ in the gravitational action. This model, known as Higgs inflation [77], shares a lot of similarities with Starobinsky's original proposal, which extends the gravitational action by an $\mathcal{R}^2$ term [70]. Currently, the predictions of Higgs and $\mathcal{R}^2$ inflation are in excellent agreement with the most recent CMB data. Other field-theoretical models attempt to embed inflation into grand unification or establish a connection with the accelerated expansion of the current Universe. The latter scenario is also known as quintessential inflation because it identifies the inflaton and quintessence as the same scalar field [78]. This represents an economical scenario, even though there is no empirical reason to believe that the accelerated cosmic expansion at early and late times must be caused by the same dynamics.

Ultimately, inflation requires a quantum description of gravity, in order to maintain control over gravitational corrections to the potential energy density of the inflaton field. This motivates the construction of inflation models in the context of string

theory [79], which is one the main ambitions of the field of string cosmology [80]. Inflation may even offer the possibility to test predictions of string theory by means of cosmological observations, which is otherwise a notoriously difficult task. The vast landscape of possible vacuum states in string theory moreover lends support to the notion of the multiverse, i.e., the speculative idea that our Universe is just one among countless other universes that are continuously born during inflation, each possibly in a different vacuum state. Anthropic arguments applied to inflation in the string landscape may notably explain the tiny value of the cosmological constant [81], albeit such inferences, which most likely elude any possible means of experimental confirmation, are a source of controversy. On the other hand, one may regard the multiverse as an unavoidable consequence of any stage of inflation that globally lasts forever but locally decays into causally disconnected bubbles. Inflation of this type is known as *eternal inflation* and a generic prediction of many models [82].

Much of the current quantum-gravity research on inflation focuses on its feasibility in view of the so-called *swampland conjectures*, which attempt to delineate the boundary between effective field theories that do and that do not possess a consistent ultraviolet completion in quantum gravity, i.e., theories in the "landscape" and the "swampland" [83]. Several of these conjectures pose a severe challenge to simple models of inflation [84, 85], meaning that more sophistication may be needed to construct fully realistic models. The same holds true for the initial conditions of inflation, which often require a certain amount of fine-tuning, depending on the speculative assumptions about the pre-inflationary era. This is of course undesirable as the entire *raison d'être* of inflation is to resolve the fine-tuning issues related to the initial conditions of big-bang cosmology. For some authors, these shortcomings of inflation serve as a motivation to consider some of its (arguably less popular) extensions or alternatives, such as as big-bounce models [86], cyclic models [87], or string gas cosmology [88]. Most authors, however, adhere to the paradigm of inflationary cosmology, remaining confident that future work in field theory, string theory, and quantum cosmology [89], in combination with more data, will resolve the outstanding issues.

3.2 Hot Thermal Phase

Inflation ends whenever the equation of state of the inflaton field (or fields) no longer supports an accelerating expansion. This happens, e.g., when the inflaton field picks up a large kinetic energy, i.e., when the so-called slow-roll conditions become violated, or when a critical field value triggers a phase transition in scalar field space. In the latter case, the phase transition must be sufficiently smooth; a first-order phase transition followed by bubble collisions as in early models of inflation would reintroduce unacceptably large inhomogeneities. In other words, inflation must allow for a *graceful exit* into the stage of decelerating expansion, an exit that does not spoil the homogeneity and isotropy established by inflation. During this exit, the vacuum energy stored in the inflaton field is converted to thermal radiation

composed of relativistic particles. In this sense, assuming that inflation was preceded by a primordial radiation-dominated era at extremely high temperatures, one can say that the decay of the inflaton field after inflation *reheats* the Universe. However, even independently of any assumptions about the pre-inflationary era, the process of entropy production at the end of inflation is nowadays commonly referred to as *reheating* [90]. Correspondingly, the temperature of the thermal bath at the end of reheating, when the radiation energy density begins to dominate the cosmic energy budget, is known as the reheating temperature $T_{\rm rh}$. This temperature marks the onset of the hot big bang and is an important quantity in the description of the early Universe. At present, its value is only weakly constrained, $T_{\rm rh} \sim \text{few} \times 10^{-3} \cdots 10^{15}\,\text{GeV}$, which leaves room for countless possible scenarios of the evolution of the early Universe between the end of inflation and BBN [91]. In the following, we will, however, ignore the possibility of a nonstandard expansion history and focus on the most important events during the standard radiation-dominated era after reheating; see Fig. 4.

Many models predict a very large reheating temperature, far beyond the energy scales that are accessible in terrestrial collider experiments. In this sense, the hot thermal plasma in the early Universe represents a unique particle physics laboratory that provides the right environment for particle processes at extremely high energies. This observation serves as the starting point for the field of particle cosmology [92], which applies the methods of theoretical particle physics to the early Universe in order to better understand the phenomenology of the Standard Model as well as of hypothetical new-physics scenarios at very high temperatures.

A central subject of particle cosmology is the evolution of the subatomic forces in the early Universe, which is related to the phenomenon of spontaneous symmetry breaking in the context of gauge theories. At temperatures of a few 100 GeV, the gauge symmetry governing the interactions of elementary particles is given by the SM gauge group, $G_{\rm SM} = SU(3)_C \times SU(2)_L \times U(1)_Y$, which describes the strong and electroweak gauge interactions in the bath of massless SM particles. The origin of the SM gauge structure is unknown; but there are good reasons to believe that $G_{\rm SM}$ is in fact just a subgroup of a much larger symmetry group $G_{\rm GUT}$ that unifies all subatomic forces at an energy scale of around 10^{16} GeV, a hypothesis that goes by the name of *grand unification*. In each grand unified theory, the gauge groups $G_{\rm GUT}$ and $G_{\rm SM}$ are connected by a characteristic pattern of symmetry breaking steps, some of which may take place before inflation (e.g., in order to avoid a monopole problem; see Sect. 3.1) and some of which may take place after inflation. Similarly, the electroweak part of the SM gauge group itself, $SU(2)_L \times U(1)_Y$, becomes spontaneously broken at temperatures of around 160 GeV in consequence of the SM Higgs mechanism. This mechanism manifests itself in a smooth cosmological phase transition, the electroweak crossover, that yields a nonzero vacuum expectation value of the Higgs field, which in turn induces masses for all SM particles (possibly except for neutrinos whose masses call for new physics).

After the electroweak phase transition, the temperature keeps decreasing. Heavy degrees of freedom, such as the top quark or electroweak gauge bosons, are therefore no longer produced in the thermal bath, while the majority of the remaining

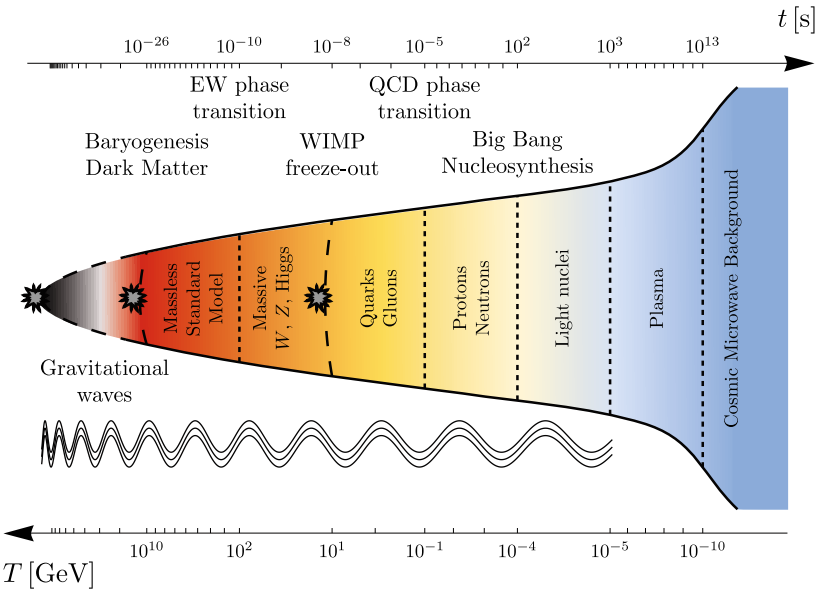

Fig. 4 Chronology of the hot big bang, i.e., the radiation-dominated era after inflation and reheating [93]. Three possible values of the maximal temperature in the early Universe are indicated by little gray stars. Different epochs are labeled by their respective characteristic degrees of freedom and separated by vertical dashed lines. Gravitational waves produced at very early times can freely propagate through the hot plasma, while photons only decouple from the thermal bath around 380,000 years after the big bang, shortly after matter–radiation equality. During radiation domination, the scale factor grows like $R \propto t^{1/2}$, while it scales like $R \propto t^{2/3}$ during matter domination, as indicated by the growing width of the black solid envelope. Image credit: The author

light degrees of freedom contribute to the so-called quark–gluon plasma, the state of strongly interacting matter in the high-temperature limit of quantum chromodynamics (QCD). At temperatures of around 160 MeV, the plasma of quarks and gluons become confined in color-neutral hadrons. This process manifests itself again in a smooth phase transition, the QCD crossover or simply quark–hadron phase transition. Shortly thereafter, baryons and antibaryons begin to decouple, which results in the annihilation of more or less all antibaryons and the freeze-out of a small relic baryon density (see Sect. 2.1). While this process takes place only after the QCD phase transitions, the primordial asymmetry between baryons and antibaryons was presumably already seeded much earlier. Standard thermal leptogenesis, e.g., one of the most popular baryogenesis scenarios, operates at temperatures of the order of 10^9 GeV or even higher.

Around one second into the radiation-dominated era, the thermal bath consists of relativistic photons, neutrinos, electrons, and positrons as well as nonrelativistic protons and neutrons. At this point in time, neutrinos participate for the last time in scattering processes mediated by the weak nuclear force, before they decouple and begin to free-stream in the form of the cosmic neutrino background (CνB). Like the

CMB, the CνB is a firm prediction of big-bang cosmology. Because of the elusive nature of neutrinos, it has not yet been observed [94], although experiments such as PTOLEMY [95] aim at its direct detection in the not-too-distant future. Shortly after neutrino decoupling, at temperatures around the electron mass scale, electrons and positrons annihilate into photons. This leaves behind a relic electron density (which is tied to the relic proton density by the requirement that the Universe must remain charge-neutral) and slightly heats up the photon bath. As a consequence, the CMB is expected to have a slightly larger temperature than the CνB. Or conversely, while the CMB temperature today is known to be 2.73 K, the CνB is predicted to have a temperature of only 1.95 K.

After two minutes or so, primordial nucleosynthesis commences. The temperature has now dropped to around 100 keV, which is low enough so that the formation of stable deuterium is no longer impeded by photodissociation. Following the passage of this so-called deuterium bottleneck, a network of nuclear reactions unfolds, which mostly leads to the synthesis of deuterium, helium-3, helium-4, and lithium-7 (see Sect. 1.3). In particular, all neutrons become bound in atomic nuclei, with the vast majority of neutrons ending up in helium-4. After around 20 min, BBN concludes, which determines the characteristic composition of the plasma in terms of photons, electrons, protons, and light nuclei that it will keep for the entire remainder of the radiation-dominated era.

The end of the radiation-dominated era is reached around 50,000 years after the big bang, when the relic density of dark matter (see Sect. 2.2) begins to dominate the energy budget of the Universe. Dark matter is presumably produced at very high temperatures, maybe already during reheating after inflation. In a popular class of models, dark matter consists, e.g., of weakly interacting massive particles (WIMPs), particles with a mass of the order of 100 GeV and an interaction strength comparable to the strength of electroweak interactions [96]. Such particles can be thermally produced in the early Universe, before their relic density freezes out at temperatures that are about a factor 30 smaller than their mass (see Fig. 4). This is, however, only one among an almost countless number of possible scenarios for the production of dark matter in the early Universe. In any case, it is clear that, once the dark-matter particles become nonrelativistic, their energy density decreases because of the cosmic expansion in proportion to R^{-3}. That is, the energy stored in the rest mass of the dark-matter particles remains constant, while the volume that the particles are contained in grows like R^3. This needs to be compared to the dilution of the radiation energy density, which decreases in proportion to R^{-4}. Here, three powers of the scale factor account again for the volume expansion of the Universe, while the fourth power is a consequence of cosmological redshift, i.e., the fact that the wavelength of radiation traveling through the expanding Universe becomes stretched. Because of these different scaling laws, the energy density of matter necessarily begins to dominate over the energy density of radiation sooner or later. Given the relic density of dark matter today, which can be measured in CMB observations, one is then able to compute the time of matter–radiation equality in the early Universe, 50,000 years after the big bang, which marks the onset of the matter-dominated era.

Most of the events described above occur before BBN and thus fall into uncharted territory. In particular, because the early Universe is opaque to photons throughout radiation domination, no signal in the form of electromagnetic radiation will ever reach us from this era. Therefore, in order to push the frontier of our knowledge further back in time, an important line of attack necessarily consists of indirect probes, i.e., astrophysical and cosmological observations of the late Universe after photon decoupling in combination with terrestrial experiments. However, alongside these indirect approaches, there is also one direct messenger from the early Universe that can give us a peek behind the veil of the CMB: primordial gravitational waves [97], which travel more or less freely through the early Universe after their production. Many processes in the early Universe can give rise to a stochastic background of primordial gravitational waves, ranging from inflation and reheating over first-order cosmological phase transitions to topological defects such as domain walls and cosmic strings. The detection of such a background would undoubtedly mark a major milestone in the still young field of GW astronomy.

In fact, a new low-frequency signal has recently shown up in the data of pulsar timing arrays (PTAs), which monitor networks of pulsars across the Milky Way in order to search for nanohertz gravitational waves [98]. All big PTA collaborations, NANOGrav, PPTA, EPTA, and IPTA, have now found strong evidence for this signal [99–102], albeit it is not yet clear whether it really possesses all the properties of a genuine GW signal. More PTA data and analyses in the coming years will help clarify the situation and bring us closer to the first detection of a stochastic GW background. It is expected that this background will receive large contributions from inspiraling supermassive black-hole binaries at the center of merging galaxies. With a bit of luck, however, the data may also contain evidence for the presence of gravitational waves from the early Universe.

3.3 Structure Formation

During the radiation-dominated era, the shape of the gravitational potential is controlled by density perturbations in the radiation bath. This prevents the efficient growth of structure at early times, as the pressure of the baryon–photon fluid counteracts the gravitational pull towards overdense regions. On top of that, the fast cosmic expansion during radiation domination constantly drags all particles apart, which causes the gravitational potential to actually decay. The onset of efficient structure formation is therefore delayed to the time of matter–radiation equality, when the CDM density perturbations begin to take control. From this point on, the gravitational potential remains stable, which goes hand in hand with a more rapid accumulation of dark-matter particles in gravitational potential wells. Dark matter thus already begins to collapse under the influence of gravity, while the evolution of the baryon–photon fluid is still governed by the characteristic sound waves that later become imprinted in the CMB. Indeed, the growth of baryonic density perturbations only speeds up after photon decoupling, 380,000 years after the big bang, when the

baryons begin to fall towards the pre-existing dark-matter overdensities. Remarkably enough, the sound-wave pattern in the baryon density partially survives this process, which eventually leads to the BAO signature in the galaxy power spectrum at later times.

After photon decoupling, the CMB radiation becomes redshifted by the cosmic expansion, so that it quickly leaves the visible part of the electromagnetic spectrum. At this point, the Universe no longer contains any visible light sources, which marks the beginning of the so-called dark ages. During this epoch, most of the baryonic matter resides in neutral hydrogen, which occasionally absorbs CMB photons in 21 cm spin-flip transitions [103]. At the same time, the inhomogeneities in the baryon density continue to increase, such that structure formation transitions from the linear to the nonlinear regime on small scales (in a hierarchical bottom-up fashion; see Sect. 2.2). Roughly hundred million years into the dark ages, this leads to the formation of the first generation of stars (i.e., population III stars), followed by the formation of the first early galaxies and quasars (i.e., active galaxies with supermassive black holes at their centers) over the subsequent few hundred million years.

The neutral hydrogen partially absorbs and re-emits the ultraviolet Lyman-alpha radiation from the first stars, before it is heated by luminous X-ray sources, such as active galactic nuclei or the hot interstellar medium. This epoch is known as cosmic dawn, which results in a second characteristic 21 cm absorption feature in the CMB background. In 2018, the EDGES collaboration announced the first detection of this signal with an unexpectedly large amplitude [104]; however, since then, no other experiment has been able to confirm this claim [105]. Future radio observations by LOFAR [106, 107]/NenuFAR [108] and the Square Kilometre Array (SKA) [109] will be crucial to clarify the case. Meanwhile, the epoch of cosmic dawn will also be probed by the James Webb Space Telescope, which was launched in 2021 and is expected to collect light from the first stars and galaxies [110].

Another important effect of the first stellar light in the Universe is that it gradually ionizes the neutral hydrogen gas. This process is referred to as reionization [111], lasts from around 250 million years to around one billion years after the big bang, and is powered by the energetic ultraviolet radiation emitted by the first stars, galaxies, and quasars, which separates again electrons from protons and hence reverts the outcome of recombination. Direct evidence for the epoch of reionization is, e.g., provided by quasar absorption spectra. High-redshift quasars feature broadened and redshifted Lyman-alpha absorption lines of neutral hydrogen [112, 113], known as Gunn–Peterson troughs [114], while the spectra of quasars located at lower redshifts exhibit no such troughs. Based on these observations, one concludes that the intergalactic medium must have been neutral at early times, before it became fully ionized towards the end of the first billion years. After reionization, neutral hydrogen only remains in isolated gas clouds, which results in a sequence of individual Lyman-alpha absorption lines in the spectra of distant quasars. This so-called Lyman-alpha forest can be used as a tracer of the large-scale structure of the Universe and hence as a tool to constrain cosmological parameters [115]. A closely related probe of the large-scale structure consists of hydrogen intensity mapping, i.e., the measurement of 21 cm emission from

unresolved clouds of neutral hydrogen [116]. The promise of this technique is that it will eventually enable us to map the three-dimensional matter distribution across vast spatial volumes. In a first step towards this goal, the CHIME experiment recently reported the detection of 21 cm emission from select galaxies and quasars [117]; next, it aims at measuring the BAO signature in the auto correlation power spectrum of 21 cm emission.

Structure formation continues throughout and after reionization, which gradually results in what is known as the *cosmic web*, the familiar distribution of matter that shapes the appearance of the Universe up to this day. The cosmic web consists of massive dark-matter halos, connected to each other by filamentary structures and surrounded by vast underdense voids. In the cosmic web, the baryonic matter traces the distribution of cold dark matter, which results in the formation of galaxies in overdense regions. Over time, these galaxies assemble in galaxy clusters, with the most massive superclusters forming at the center of dense dark-matter halos. In this process, the baryonic matter can condense into complex structures even on very small scales, because, unlike dark matter, it is able to cool and loose angular momentum via the emission of electromagnetic radiation.

The dynamics and evolution of the cosmic web is a highly nonlinear process that is best studied in numerical N-body simulations (see Fig. 5), an important tool of modern cosmology. State-of-the-art N-body simulations are very successful in describing the evolution of structure on large length scales. On shorter length scales, however, they currently face a number of challenges that demand further scrutiny [118]. One of these small-scale problems consists, e.g., in the observation that most dwarf galaxies appear to host a dark-matter core of roughly constant density at their center, while simulations based on the ΛCDM model rather indicate a cuspy density profile. Other

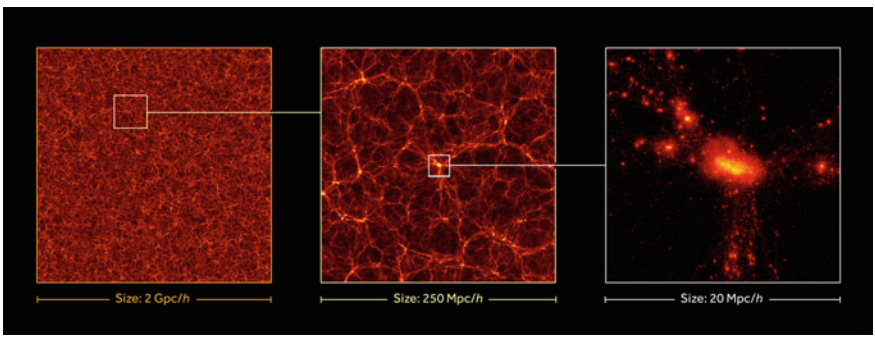

Fig. 5 Visualization of the AbacusSummit simulation [124], a state-of-the-art cosmological N-body simulation of roughly 330 billion test particles inside a volume of $(2\,\mathrm{Gpc}/h)^3$, where $h \approx 0.7$ is a dimensionless measure for the Hubble constant H_0. On gigaparsec scales (left), the matter distribution is homogeneous and isotropic in accord with the cosmological principle; on scales of around 100 megaparsec (middle), the cosmic web consisting of halos, filaments, and voids becomes apparent; on scales of around 10 megaparsec (right), individual halos and galaxy clusters are resolved. Image credit: The AbacusSummit Team

problems, next to this cusp–core problem, are known as the missing-satellites problem or the too-big-to-fail problem. The origin of these small-scale problems might be related to the complicated baryonic feedback on structure formation on small scales, which is currently not well understood, or point to nonstandard properties of dark matter, such as weak dark-matter self-interactions or a warm dark-matter component. Progress at this frontier will require improved theoretical and numerical modelling as well as more astronomical data, such as, e.g., the Legacy Survey of Space and Time (LSST) conducted by the Vera C. Rubin Observatory [119].

Our own galaxy, the Milky Way, forms around five billion years after the big bang, followed by the solar system roughly four billion years later. Another billion years later, around ten billion years after the big bang, dark energy begins to dominate the cosmic energy budget, which marks the onset of the late-time accelerated expansion (see Sect. 2.3). This transition from matter domination to dark-energy domination is a consequence of the fact that the energy density of matter steadily decreases because of the volume expansion, $\rho_m \propto R^{-3}$, while the energy density of dark energy remains constant, $\rho_\Lambda \propto R^0$. The cosmic expansion thus constantly increases the absolute amount of dark energy, which may appear counterintuitive from the perspective of classical mechanics, but which is indeed a perfectly viable possibility in the context of general relativity in an expanding Universe. An important effect of dark energy on structure formation is that it causes the gravitational potential to decay again. In the late Universe, the growth of structure because of gravitational instability can no longer compete with the accelerated expansion, such that wells and hills in the gravitational potential begin to decay. CMB photons traveling through this decaying potential landscape will therefore pick up small additional gravitational red- and blueshifts. This phenomenon is known as the late-time integrated Sachs–Wolfe effect [120], or Rees–Sciama effect [121], and has been observed in the cross correlation of the CMB with the large-scale structure [122, 123].

In the coming billion years, our Universe will continue to appear similar to its present state, albeit the accelerated expansion will continue to push more and more galaxies beyond our cosmic horizon. Thus, assuming that the accelerated expansion continues far into the future, the observable Universe will eventually only contain the gravitationally bound group of galaxies in our immediate cosmic neighbourhood; and these galaxies will gradually grow dimmer over time as the fuel for the formation of new stars becomes depleted. This extrapolation, however, depends on our assumptions about the nature of dark energy and therefore needs to be treated with caution. Statements about the ultimate fate of the Universe are even more speculative, whereby we have come full circle and returned to the starting point of our chronology of the Universe in this section.

4 Status and Prospects

In the course of the hundred years since Einstein and Hubble, cosmology has advanced from a subdiscipline of astronomy to a cutting-edge research field in its

own right, closely interwoven with other strands of basic science. Today, cosmology is an interdisciplinary and international endeavor that epitomizes the human quest for knowledge, humanity's desire to better understand the cosmos and its place in it. To this end, cosmology brings together insights from many neighbouring fields, ranging from string theory, field theory and high-energy physics over neutrino physics, astroparticle physics, and astronomy to nuclear, atomic, and gravitational physics. Modern cosmological observations and simulations are in particular a prime incarnation of big-data science and benefit as such from the steady advances in computing power and new algorithms, e.g., based on machine learning [125–127]. In addition, the diverse observational program of modern cosmology is a shining example of collaboration science that capitalizes on community practices such as queue observing and open-access archives.

The progress in the field over the last century and especially over the last decades has been impressive. Based on several major breakthrough detections, including the Hubble expansion, the CMB and its anisotropies as well as dark energy, cosmology now possesses a rigorously formulated standard model, the ΛCDM concordance model, which manages to explain the bulk of all cosmological data with great precision in terms of just a few basic building blocks: ordinary baryonic matter, cold nonbaryonic dark matter, and a cosmological constant. We now understand that matter density perturbations, probably seeded by an early stage of cosmic inflation before the hot big bang, gave rise to the large-scale structure of the Universe, while the cosmological constant completes the cosmic energy budget required for a flat Universe, which causes the present-day expansion to accelerate. Since the COBE measurement of the CMB blackbody spectrum in the early 1990s, we have moreover entered the era of precision cosmology [128]. A large number of cosmological parameters are now known with percent-level precision, which is a testament to the long way that modern cosmology has come since the days of Hubble, whose determination of the Hubble constant, $H_0 \approx 500\,\mathrm{km/s/Mpc}$, was not even of the correct order of magnitude.

Despite its overall success, the ΛCDM model as well as the precise determination of its parameters leave us with profound questions that remain to be answered. First of all, it is remarkable that all of the basic ΛCDM building blocks require new physics beyond the Standard Model of particle physics: The Standard Model does not contain a viable particle candidate for dark matter; naive QFT estimates of the energy density of the vacuum exceed the observed value of the cosmological constant by an embarrassingly large amount; and the cosmic baryon–antibaryon asymmetry calls for an exotic nonequilibrium process in the early Universe. At the same time, big-bang cosmology, while successful in explaining the primordial abundances of the light elements as well as the properties of the CMB, lacks a satisfying explanation of its initial conditions. The current paradigm for the dynamics underlying these initial conditions is cosmic inflation, which represents yet another phenomenon that cannot be explained in terms of SM physics in combination with Einstein gravity. Together, these open questions are key drivers of the research program of particle physics and cosmology in the 21st century.

More challenges arise from the precise measurement of cosmological parameters, which put the ΛCDM framework itself to the test. Now, in the era of precision cosmology, different routes to measuring the same observable can especially be used to set up powerful consistency checks that may reveal even slight discrepancies. Indeed, in the recent past, various such tensions have emerged, some more subtle and others more severe [24]. The list of tensions, challenges, and oddities includes (i) various CMB anomalies [129], such as a lack of sizable temperature correlations on large angular scales or gravitational-lensing effects above the expected level; (ii) the small-scale problems of structure formation (see Sect. 3.3), such as the cusp–core or missing-satellites problems; (iii) observations challenging the cosmological principle, such as dipole measurements, e.g., in the distribution of quasars [23], that do not match the properties of the CMB dipole; and several others. On top of that, there is an on-going debate related to different measurements of the clustering amplitude σ_8, an observable that characterizes the size of matter density fluctuations in the late Universe (in the linear-theory approximation) on length scales of the order of $8\,\mathrm{Mpc}/h$, i.e., on galaxy-cluster scales (see Fig. 5). Some weak-lensing and galaxy-clustering surveys find σ_8 values below the value inferred by the PLANCK observations of the CMB, at a statistical significance of up to roughly $3\,\sigma$ [130, 131], while other surveys find values closer to the CMB value [132]. It therefore remains to be seen whether the σ_8 discrepancy will eventually reveal a crack in the ΛCMB paradigm or not.

The elephant in the room, however, dwarfing all of the anomalies listed above, is the so-called H_0 or Hubble tension [133, 134], a mismatch between different determinations of the Hubble constant that has recently reached a statistical significance of $5\,\sigma$ [135] and which has even led some commentators to proclaim a *crisis in cosmology*. Of course, such a proclamation is overly dramatic and in a sense also misleading: If anything, the Hubble tension illustrates that cosmology as a scientific discipline is in fact in a very healthy state. It has now become a mature discipline of precision science, where a steady influx of new data enables one to make real and meaningful progress. In the case of the Hubble tension, this state of affairs is reflected in increasingly precise determinations of the Hubble constant. On the one hand, the "global" measurement of H_0 based on CMB observations by the PLANCK satellite, assuming a standard ΛCDM cosmology, yields a value of $H_0 = 67.4 \pm 0.5\,\mathrm{km/s/Mpc}$ [33]. On the other hand, recent "local" measurements at low redshift based on the observation of Cepheid variables and Type Ia supernovae (i.e., the *cosmic distance ladder*) consistently find larger values, $H_0 = 73.0 \pm 1.0\,\mathrm{km/s/Mpc}$ [135]. The origin of this discrepancy is unknown, which currently fuels intense research efforts in particle physics and cosmology aimed at identifying possible solutions. Next to systematic effects, which may account for parts of the tension, various possibilities are presently scrutinized, including departures from ΛCDM in the early Universe, departures from ΛCDM in the late Universe, as well as combinations thereof [136]. Many proposed solutions also promise other observational signatures that may eventually enable one to identify the correct solution. However, when and in which specific way this is going to happen is impossible to predict at the moment.

In any case, it is clear that cosmology is facing a bright future in the coming years and decades, regardless of whether the Hubble tension will still stay with us for a bit longer or not. A comprehensive outlook is beyond the scope of this essay; however, a few exciting prospects deserve to be highlighted. On the observational side, a wealth of data from next-generation space telescopes is awaiting us, including NASA's James Webb [110] and Nancy Grace Roman [137] as well as ESA's Euclid [138] and Athena [139] space missions. These space telescopes will be complemented by new optical reflecting telescopes on the ground, such as the Vera C. Rubin Observatory [119], the Thirty Meter Telescope [140], and the Extremely Large Telescope, alongside the next generation of radio and gamma-ray telescopes, SKA [109] and the imaging atmospheric Cherenkov Telescope Array (CTA) [141], as well as a suite of new space- and ground-based CMB experiments (see Sect. 1.2). These new telescopes and observatories will enable a revolutionary research program of extraordinary depth and breadth, which will take proven observational methods (e.g., galaxy surveys, weak gravitational lensing, supernova cosmology) to new heights and establish novel techniques that have currently not yet reached their full potential (e.g., cosmic chronometers, gamma-ray bursts [142], 21 cm cosmology) as new tools in the cosmologist's toolbox [143, 144].

Moreover, we can expect groundbreaking developments at the multi-messenger frontier, i.e., in the fields of neutrino and GW cosmology. In particular, the third generation of ground-based GW interferometers, Cosmic Explorer [145] and Einstein Telescope [146], as well as the first GW space missions, the Laser Interferometer Space Antenna (LISA) [147], Taiji [148], and TianQin [149], promise to open new windows onto the Universe. GW events can, e.g., be used as *standard sirens* [150], in analogy to type-Ia supernova *standard candles*, in order to measure cosmological parameters. Furthermore, the detection of a cosmological GW background would grant us access to processes in the very early Universe that are currently hidden behind the veil of the CMB. In fact, the current generation of PTA observations may actually be on the brink of detecting a GW background at nanohertz frequencies (see Sect. 3.2). While this signal is expected to be dominated by the astrophysical foreground from supermassive black-hole binaries, there is a slight chance that it may also contain a cosmological component.

Finally, the rich observational program in the coming years will also stimulate and profit from theoretical progress, including the development of new models, methods, and algorithms. We thus find ourselves on the eve of a golden age of cosmology that will likely shape our understanding of the cosmos in ways that we can barely imagine at present. Along the way, we may gradually encounter solutions to the various challenges and tensions of the ΛCDM model, which may or may not usher in the next paradigm shift in cosmology. Either way, a fascinating journey lies ahead of us that is set to resolve various deep questions and give rise to new ones. The first century of modern cosmology has been a true adventure, the second century will be even more exciting.

References

1. E.W. Kolb, M.S. Turner, *The Early Universe* (CRC Press, Boca Raton, 1994)
2. V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, Cambridge, 2005)
3. S. Weinberg, *Cosmology* (Oxford University Press, Oxford, 2008)
4. S. Dodelson, F. Schmidt, *Modern Cosmology*, 2nd edn. (Academic, Amsterdam, 2020)
5. P.J.E. Peebles, *Cosmology's Century* (Princeton University Press, Princeton, 2020)
6. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973)
7. S.M. Carroll, *Spacetime and Geometry* (Cambridge University Press, Cambridge, 2019)
8. A. Friedmann, Über die Krümmung des Raumes. Zeitschrift für Physik **10**, 377 (1922). https://doi.org/10.1007/BF01332580
9. R. Barnes, The Friedman-Lemaître-Robertson-Walker Metric: a centennial review, arXiv:2201.13120
10. G. Lemaître, A homogeneous universe of constant mass and growing radius accounting for the radial velocity of extragalactic nebulae. Annales Soc. Sci. Bruxelles A **47**, 49 (1927). https://doi.org/10.1007/s10714-013-1548-3
11. E. Hubble, A relation between distance and radial velocity among extra-galactic nebulae. Proc. Nat. Acad. Sci **15**, 168 (1929). https://doi.org/10.1073/pnas.15.3.168
12. E. Hubble, M.L. Humason, The velocity-distance relation among extra-galactic nebulae. Astrophys. J **74**, 43 (1931). https://doi.org/10.1086/143323
13. V. Slipher, The radial velocity of the Andromeda Nebula. Lowell Obs. Bull. **1**, 56 (1913)
14. K. Lundmark, The determination of the curvature of space-time in de Sitter's world. Mon. Not. Roy. Astron. Soc. **84**, 747 (1924)
15. E.R. Harrison, *Darkness at Night: A Riddle of the Universe* (Harvard University Press, Cambridge, 1989)
16. G. Gamow, The evolution of the Universe. Nature **162**, 680 (1948). https://doi.org/10.1038/162680a0
17. R.A. Alpher, R.C. Herman, Evolution of the Universe. Nature **162**, 774 (1948). https://doi.org/10.1038/162774b0
18. R.A. Alpher, R.C. Herman, On the relative abundance of the elements. Phys. Rev **74**, 1737 (1948). https://doi.org/10.1103/physrev.74.1737
19. P.J.E. Peebles, Recombination of the primeval plasma. Astrophys. J **153**, 1 (1968). https://doi.org/10.1086/149628
20. R. Durrer, *The Cosmic Microwave Background* (Cambridge University Press, Cambridge, 2020)
21. A.A. Penzias, R.W. Wilson, A Measurement of excess antenna temperature at 4080-Mc/. Astrophys. J **142**, 419 (1965). https://doi.org/10.1086/148307
22. D.J. Fixsen, E.S. Cheng, J.M. Gales, J.C. Mather, R.A. Shafer, E.L. Wright, The Cosmic Microwave Background spectrum from the full COBE FIRAS data set. Astrophys. J **473**, 576 (1996). https://doi.org/10.1086/178173, arXiv:astro-ph/9605054
23. N.J. Secrest, S. von Hausegger, M. Rameez, R. Mohayaee, S. Sarkar, J. Colin, A test of the cosmological principle with quasars. Astrophys. J. Lett **908**, L51 (2021). https://doi.org/10.3847/2041-8213/abdd40, arXiv:2009.14826
24. L. Perivolaropoulos, F. Skara, *Challenges for ΛCDM: An update*, arXiv:2105.05208
25. WMAP collaboration, Nine-year Wilkinson microwave anisotropy probe (WMAP) observations: final maps and results. Astrophys. J. Suppl **208**, 20 (2013). https://doi.org/10.1088/0067-0049/208/2/20, arXiv:1212.5225
26. Planck collaboration, Planck 2018 results. I. Overview and the cosmological legacy of Planck. Astron. Astrophys. **641**, A1 (2020) https://doi.org/10.1051/0004-6361/201833880. arXiv:1807.06205
27. H. Hui et al., BICEP Array: a multi-frequency degree-scale CMB polarimeter. https://doi.org/10.1117/12.2311725 Proc. SPIE Int. Soc. Opt. Eng**10708**, 1070807 (2018). arXiv:1808.00568

28. CMB-S4 collaboration, *CMB-S4 Science Book, First Edition*. arXiv:1610.02743
29. LiteBIRD collaboration, *Probing Cosmic Inflation with the LiteBIRD Cosmic Microwave Background Polarization Survey*. arXiv:2202.02773
30. NASA PICO collaboration, *PICO: Probe of Inflation and Cosmic Origins*. arXiv:1902.10541
31. Simons Observatory collaboration, The Simons Observatory: Science goals and forecasts. JCAP **02**, 056 (2019). https://doi.org/10.1088/1475-7516/2019/02/056, arXiv:1808.07445
32. D.J. Fixsen, The Temperature of the Cosmic Microwave Background. Astrophys. J. **707**, 916 (2009). https://doi.org/10.1088/0004-637X/707/2/916 arXiv: 0911.1955
33. Planck collaboration, Planck 2018 results. VI. Cosmological parameters. Astron. Astrophys. **641**, A6 (2020). https://doi.org/10.1051/0004-6361/201833910, arXiv:1807.06209
34. R.A. Alpher, H. Bethe, G. Gamow, The origin of chemical elements. Phys. Rev. **73**, 803 (1948). https://doi.org/10.1103/PhysRev.73.803
35. R.H. Cyburt, B.D. Fields, K.A. Olive, T.-H. Yeh, Big Bang Nucleosynthesis: 2015. Rev. Mod. Phys. **88**, 015004 (2016). https://doi.org/10.1103/RevModPhys.88.015004, arXiv: 1505.01076
36. C. Pitrou, A. Coc, J.-P. Uzan, E. Vangioni, Precision big bang nucleosynthesis with improved Helium-4 predictions. Phys. Rept. **754**,1 (2018). https://doi.org/10.1016/j.physrep.2018.04.005, arXiv:1801.08023
37. B.D. Fields, The primordial lithium problem. Ann. Rev. Nucl. Part. Sci. **61**, 47 (2011). https://doi.org/10.1146/annurev-nucl-102010-130445, arXiv: 1203.3551
38. V. Mossa et al., The baryon density of the Universe from an improved rate of deuterium burning. Nature **587**, 210 (2020). https://doi.org/10.1038/s41586-020-2878-4
39. Particle Data Group collaboration, Review of Particle Physics. PTEP **2020**, 083C01 (2020). https://doi.org/10.1093/ptep/ptaa104
40. M. Archidiacono, E. Giusarma, S. Hannestad, O. Mena, Cosmic dark radiation and neutrinos. Adv. High Energy Phys. **2013**, 191047 (2013). https://doi.org/10.1155/2013/191047, arXiv: 1307.0637
41. M. Kawasaki, K. Kohri, T. Moroi, Big-Bang nucleosynthesis and hadronic decay of long-lived massive particles. Phys. Rev. D **71**, 083502 (2005). https://doi.org/10.1103/PhysRevD.71.083502 arXiv: astro-ph/0408426
42. W.J. Percival, S. Cole, D.J. Eisenstein, R.C. Nichol, J.A. Peacock, A.C. Pope et al., Measuring the Baryon Acoustic Oscillation scale using the SDSS and 2dFGRS. Mon. Not. Roy. Astron. Soc. **381**, 1053 (2007). https://doi.org/10.1111/j.1365-2966.2007.12268.x arXiv: 0705.3323
43. SDSS collaboration, Baryon Acoustic Oscillations in the Sloan Digital Sky Survey Data Release 7 Galaxy Sample. Mon. Not. Roy. Astron. Soc. **401**, 2148 (2010). https://doi.org/10.1111/j.1365-2966.2009.15812.x, arXiv:0907.1660
44. Supernova Search Team collaboration, Observational evidence from supernovae for an accelerating universe and a cosmological constant. Astron. J. **116**, 1009 (1998). https://doi.org/10.1086/300499, arXiv: astro-ph/9805201
45. Supernova Cosmology Project collaboration, Measurements of Ω and Λ from 42 high redshift supernovae. Astrophys. J **517**, 565 (1999). https://doi.org/10.1086/307221, arXiv: astro-ph/9812133
46. eBOSS collaboration, Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory. Phys. Rev. D **103**, 083533 (2021). https://doi.org/10.1103/PhysRevD.103.083533, arXiv: 2007.08991
47. A.D. Sakharov, Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. Pisma Zh. Eksp. Teor. Fiz **5**, 32 (1967). https://doi.org/10.1070/PU1991v034n05ABEH002497
48. D. Bodeker, W. Buchmuller, Baryogenesis from the weak scale to the grand unification scale. Rev. Mod. Phys **93**, 035004 (2021). https://doi.org/10.1103/RevModPhys.93.035004, arXiv: 2009.07294
49. P. Di Bari, On the origin of matter in the Universe. Prog. Part. Nucl. Phys. **122**, 103913 (2022) https://doi.org/10.1016/j.ppnp.2021.103913, arXiv: 2107.13750

50. M. Fukugita, T. Yanagida, Baryogenesis Without Grand Unification. Phys. Lett. B **174**, 45 (1986). https://doi.org/10.1016/0370-2693(86)91126-3
51. W. Buchmuller, R.D. Peccei, T. Yanagida, Leptogenesis as the origin of matter. Ann. Rev. Nucl. Part. Sci **55**, 311 (2005). https://doi.org/10.1146/annurev.nucl.55.090704.151558, arXiv: hep-ph/0502169
52. G. Bertone, D. Hooper, J. Silk, Particle dark matter: Evidence, candidates and constraints. Phys. Rept **405**, 279 (2005). https://doi.org/10.1016/j.physrep.2004.08.031, arXiv: hep-ph/0404175
53. J. Silk et al., *Particle Dark Matter: Observations* (Models and Searches. Cambridge University Press, Cambridge, 2010)
54. M. Milgrom, A Modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. Astrophys. J **270**, 365 (1983). https://doi.org/10.1086/161130
55. F. Zwicky, Die Rotverschiebung von extragalaktischen Nebeln. Helv. Phys. Acta **6**, 110 (1933). https://doi.org/10.1007/s10714-008-0707-4
56. F. Zwicky, On the Masses of Nebulae and of Clusters of Nebulae. Astrophys. J **86**, 217 (1937). https://doi.org/10.1086/143864
57. V.C. Rubin, W.K. Ford, Jr., Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. Astrophys. J **159**, 379 (1970). https://doi.org/10.1086/150317
58. V.C. Rubin, N. Thonnard, W.K. Ford, Jr., Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. Astrophys. J **238**, 471 (1980). https://doi.org/10.1086/158003
59. W. Hu, R. Barkana, A. Gruzinov, Cold and fuzzy dark matter, Phys. Rev. Lett **85**, 1158 (2000). https://doi.org/10.1103/PhysRevLett.85.1158 arXiv: astro-ph/0003365
60. B. Carr, F. Kuhnel, M. Sandstad, Primordial Black Holes as Dark Matter. Phys. Rev. D **94**, 083504 (2016). https://doi.org/10.1103/PhysRevD.94.083504, arXiv: 1607.06077v4
61. S. Bird, I. Cholis, J.B. Muñoz, Y. Ali-Haïmoud, M. Kamionkowski, E.D. Kovetz et al., Did LIGO detect dark matter?. Phys. Rev. Lett **116**, 201301 (2016). https://doi.org/10.1103/PhysRevLett.116.201301, arXiv: 1603.00464
62. D. Clowe, M. Bradac, A.H. Gonzalez, M. Markevitch, S.W. Randall, C. Jones et al., A direct empirical proof of the existence of dark matter. Astrophys. J. Lett. **648**, L109 (2006). https://doi.org/10.1086/508162 arXiv:astro-ph/0608407
63. K. Bamba, S. Capozziello, S. Nojiri and S.D. Odintsov, Dark energy cosmology: the equivalent description via different theoretical models and cosmography tests. Astrophys. Space Sci. **342**, 155 (2012). https://doi.org/10.1007/s10509-012-1181-8, arXiv: 1205.3421
64. D. Huterer, D.L. Shafer, Dark energy two decades after: Observables, probes, consistency tests. Rept. Prog. Phys. **81**, 016901 (2018). https://doi.org/10.1088/1361-6633/aa997e, arXiv: 1709.01091
65. S. Weinberg, Anthropic Bound on the Cosmological Constant. Phys. Rev. Lett **59**, 2607 (1987). https://doi.org/10.1103/PhysRevLett.59.2607
66. S. Weinberg, The Cosmological Constant Problem. Rev. Mod. Phys **61**, 1 (1989). https://doi.org/10.1103/RevModPhys.61.1
67. H.E.S. Velten, R.F. vom Marttens, W. Zimdahl, Aspects of the cosmological "coincidence problem". Eur. Phys. J. C **74**, 3160 (2014). https://doi.org/10.1140/epjc/s10052-014-3160-4 arXiv: 1410.2509
68. Y. Minami, E. Komatsu, New Extraction of the Cosmic Birefringence from the Planck 2018 Polarization Data. Phys. Rev. Lett **125**, 221301 (2020). https://doi.org/10.1103/PhysRevLett.125.221301, arXiv:2011.11254
69. P. Diego-Palazuelos et al., Cosmic Birefringence from Planck Data Release 4. (2022) arXiv:2201.07682
70. A.A. Starobinsky, Spectrum of relict gravitational radiation and the early state of the universe. JETP Lett. **30**, 682 (1979)
71. A.H. Guth, The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. Phys. Rev. D **23**, 347 (1981). https://doi.org/10.1103/PhysRevD.23.347
72. A.D. Linde, A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems. Phys. Lett. B **108**, 389 (1982). https://doi.org/10.1016/0370-2693(82)91219-9

73. D. Baumann, Inflation, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small*, pp. 523–686 (2011). https://doi.org/10.1142/9789814327183_0010, arXiv:0907.5424
74. D.H. Lyth and A. Riotto, Particle physics models of inflation and the cosmological density perturbation. Phys. Rept. **314**, 1 (1999). https://doi.org/10.1016/S0370-1573(98)00128-8, arXiv: hep-ph/9807278
75. A. Albrecht, P.J. Steinhardt, Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking. Phys. Rev. Lett. **48**, 1220 (1982). https://doi.org/10.1103/PhysRevLett.48.1220
76. A.D. Linde, Chaotic Inflation. Phys. Lett. B **129**, 177 (1983). https://doi.org/10.1016/0370-2693(83)90837-7
77. F.L. Bezrukov and M. Shaposhnikov, The Standard Model Higgs boson as the inflaton. Phys. Lett. B **659**, 703 (2008). https://doi.org/10.1016/j.physletb.2007.11.072, arXiv: 0710.3755
78. J. de Haro, L.A. Saló, A Review of Quintessential Inflation. Galaxies **9**, 73 (2021) https://doi.org/10.3390/galaxies9040073, arXiv: 2108.11144
79. D. Baumann, L. McAllister, *Inflation and String Theory*, Cambridge Monographs on Mathematical Physics. Cambridge University Press, 5, 2015, https://doi.org/10.1017/CBO9781316105733, arXiv: 1404.2601
80. L. McAllister and E. Silverstein, String Cosmology: A Review. Gen. Rel. Grav **40**, 565 (2008). https://doi.org/10.1007/s10714-007-0556-6 arXiv: 0710.2951
81. L. Susskind, *The Anthropic landscape of string theory*. arXiv:hep-th/0302219
82. A. Vilenkin, The Birth of Inflationary Universes. Phys. Rev. D **27**, 2848 (1983). https://doi.org/10.1103/PhysRevD.27.2848
83. E. Palti, The Swampland: Introduction and Review. Fortsch. Phys **67**, 1900037 (2019) https://doi.org/10.1002/prop.201900037, arXiv: 1903.06239
84. P. Agrawal, G. Obied, P.J. Steinhardt, C. Vafa, On the Cosmological Implications of the String Swampland. Phys. Lett. B **784**, 271 (2018) https://doi.org/10.1016/j.physletb.2018.07.040 arXiv: 1806.09718
85. S.K. Garg, C. Krishnan, Bounds on Slow Roll and the de Sitter Swampland. JHEP **11**, 075 (2019). https://doi.org/10.1007/JHEP11(2019)075, arXiv: 1807.05193
86. R.H. Brandenberger, *The Matter Bounce Alternative to Inflationary Cosmology*, arXiv: 1206.4196
87. J. Khoury, B.A. Ovrut, P.J. Steinhardt, N. Turok, The Ekpyrotic universe: Colliding branes and the origin of the hot big bang. Phys. Rev. D **64**, 123522 (2001). https://doi.org/10.1103/PhysRevD.64.123522, arXiv: hep-th/0103239
88. T. Battefeld, S. Watson, String gas cosmology. Rev. Mod. Phys **78**, 435 (2006). https://doi.org/10.1103/RevModPhys.78.435, arXiv: hep-th/0510022
89. M. Bojowald, Quantum cosmology: a review. Rept. Prog. Phys **78**, 023901 (2015). https://doi.org/10.1088/0034-4885/78/2/023901, arXiv: 1501.04899
90. L. Kofman, A.D. Linde, A.A. Starobinsky, Towards the theory of reheating after inflation. Phys. Rev. D**56**, 3258 (1997). https://doi.org/10.1103/PhysRevD.56.3258, arXiv: hep-ph/9704452
91. R. Allahverdi et al., *The First Three Seconds: a Review of Possible Expansion Histories of the Early Universe*, arXiv:2006.16182
92. Y. Mambrini, *Particles in the Dark Universe: A Student's Guide to Particle Physics and Cosmology* (Springer, Cham, 2021)
93. K. Schmitz, *The $B-L$ Phase Transition: Implications for Cosmology and Neutrinos*, Ph.D. thesis, Hamburg U., (2012). arXiv: 1307.3887
94. KATRIN collaboration, *New Constraint on the Local Relic Neutrino Background Overdensity with the First KATRIN Data Runs*, arXiv: 2202.04587
95. PTOLEMY collaboration, Neutrino physics with the PTOLEMY project: active neutrino properties and the light sterile case. JCAP **07**, 047 (2019). https://doi.org/10.1088/1475-7516/2019/07/047, arXiv: 1902.05508
96. G. Jungman, M. Kamionkowski, K. Griest, Supersymmetric dark matter. Phys. Rept **267**, 195 (1996). https://doi.org/10.1016/0370-1573(95)00058-5, arXiv: hep-ph/9506380

97. C. Caprini, D.G. Figueroa, Cosmological Backgrounds of Gravitational Waves. Class. Quant. Grav **35**, 163001 (2018). https://doi.org/10.1088/1361-6382/aac608 arXiv:1801.04268
98. S.R. Taylor, *Nanohertz Gravitational Wave Astronomy* (CRC Press, Boca Raton, 2021)
99. NANOGrav collaboration, The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background. Astrophys. J. Lett **905**, L34 (2020). https://doi.org/10.3847/2041-8213/abd401, arXiv: 2009.04496
100. B. Goncharov et al., On the Evidence for a Common-spectrum Process in the Search for the Nanohertz Gravitational-wave Background with the Parkes Pulsar Timing Array. Astrophys. J. Lett **917**, L19 (2021.) https://doi.org/10.3847/2041-8213/ac17f4, arXiv:2107.12112
101. S. Chen et al., Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: inferences in the stochastic gravitational-wave background search. Mon. Not. Roy. Astron. Soc **508**, 4970 (2021). https://doi.org/10.1093/mnras/stab2833 arXiv:2110.13184
102. 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). https://doi.org/10.1093/mnras/stab3418, arXiv: 2201.03980
103. S. Furlanetto, S.P. Oh, F. Briggs, Cosmology at Low Frequencies: The 21 cm Transition and the High-Redshift Universe. Phys. Rept **433**, 181 (2006). https://doi.org/10.1016/j.physrep.2006.08.002 arXiv:astro-ph/0608032
104. J.. Bowman, A.EE. Rogers, R.A. Monsalve, T.J. Mozdzen, N. Mahesh, An absorption profile centred at 78 megahertz in the sky-averaged spectrum. Nature **555**, 67 (2018). https://doi.org/10.1038/nature25792, arXiv: 1810.05912
105. S. Singh, J.N. T., R. Subrahmanyan, N.U. Shankar, B. S. Girish, A. Raghunathan et al., *On the detection of a cosmic dawn signal in the radio background*. arXiv:2112.06778
106. M.P. van Haarlem et al., LOFAR: The LOw-frequency ARray. Astron. Astrophys **556**, A2 (2013). https://doi.org/10.1051/0004-6361/201220873. arXiv: 1305.3550
107. T.W. Shimwell et al., *The LOFAR Two-metre Sky Survey – V. Second data release*. arXiv:2202.11733
108. F.G. Mertens, B. Semelin, L.V.E. Koopmans, Exploring the Cosmic Dawn with NenuFAR, in *Semaine de l'astrophysique française 2021*, 9, 2021, arXiv:2109.10055
109. G. Mellema et al., *Reionization and the Cosmic Dawn with the Square Kilometre Array*. Exper. Astron **36**, 235 (2013). https://doi.org/10.1007/s10686-013-9334-5, arXiv:1210.0197
110. J. P. Gardner et al., The James Webb Space Telescope. Space Sci. Rev **123**, 485 (2006). https://doi.org/10.1007/s11214-006-8315-7 arXiv:astro-ph/0606175
111. R. Barkana, A. Loeb, In the beginning: The First sources of light and the reionization of the Universe. Phys. Rept **349**, 125 (2001). https://doi.org/10.1016/S0370-1573(01)00019-9, arXiv: astro-ph/0010468
112. SDSS collaboration, Evidence for Reionization at $z \sim 6$: Detection of a Gunn-Peterson trough in a $z = 6.28$ Quasar. Astron. J **122**. 2850 (2001), https://doi.org/10.1086/324231 arXiv:astro-ph/0108097
113. X.-H. Fan, M.A. Strauss, R.H. Becker, R.L. White, J.E. Gunn, G.R. Knapp et al., Constraining the evolution of the ionizing background and the epoch of reionization with $z \sim 6$ quasars. 2. a sample of 19 quasars. Astron. J **132**, 117 (2006). https://doi.org/10.1086/504836, arXiv:astro-ph/0512082
114. J.E. Gunn, B.A. Peterson, On the Density of Neutral Hydrogen in Intergalactic Space. Astrophys. J **142**, 1633 (1965). https://doi.org/10.1086/148444
115. U. Seljak, A. Slosar, P. McDonald, Cosmological parameters from combining the Lyman-alpha forest with CMB, galaxy clustering and SN constraints. JCAP **10**, 014 (2006). https://doi.org/10.1088/1475-7516/2006/10/014 arXiv: astro-ph/0604335
116. J.R. Pritchard, A. Loeb, 21-cm cosmology. Rept. Prog. Phys **75**, 086901 (2012). https://doi.org/10.1088/0034-4885/75/8/086901 arXiv: 1109.6012
117. CHIME collaboration, *Detection of Cosmological 21 cm Emission with the Canadian Hydrogen Intensity Mapping Experiment*. arXiv: 2202.01242

118. J.S. Bullock, M. Boylan-Kolchin, Small-Scale Challenges to the Λ CDM Paradigm. Ann. Rev. Astron. Astrophys. **55**, 343 (2017). https://doi.org/10.1146/annurev-astro-091916-055313, arXiv:1707.04256
119. LSST Science, LSST Project collaboration, LSST Science Book, Version 2.0 arXiv: 0912.0201
120. R.K. Sachs, A.M. Wolfe, Perturbations of a cosmological model and angular variations of the microwave background. Astrophys. J **147**, 73 (1967). https://doi.org/10.1007/s10714-007-0448-9
121. M.J. Rees, D.W. Sciama, Large scale Density Inhomogeneities in the Universe. Nature **217**, 511 (1968). https://doi.org/10.1038/217511a0
122. P. Fosalba, E. Gaztanaga, F. Castander, Detection of the ISW and SZ effects from the CMB-galaxy correlation. Astrophys. J. Lett **597**, L8 (2003). https://doi.org/10.1086/379848 arXiv: astro-ph/0307249
123. SDSS collaboration, *Physical evidence for dark energy*, arXiv: astro-ph/0307335
124. N.A. Maksimova, L.H. Garrison, D.J. Eisenstein, B. Hadzhiyska, S. Bose, T.P. Satterthwaite, AbacusSummit: a massive set of high-accuracy, high-resolution N-body simulations. Mon. Not. Roy. Astron. Soc **508**, 4017 (2021). https://doi.org/10.1093/mnras/stab2484 arXiv: 2110.11398
125. M. Ntampaka et al., *The Role of Machine Learning in the Next Decade of Cosmology*. arXiv: 1902.10159
126. F. Villaescusa-Navarro et al., The CAMELS project: Cosmology and Astrophysics with MachinE Learning Simulations. Astrophys. J**915**, 71 (2021). https://doi.org/10.3847/1538-4357/abf7ba arXiv: 2010.00619
127. F. Villaescusa-Navarro et al., *The CAMELS project: public data release*. arXiv: 2201.01300
128. M.S. Turner, *The Road to Precision Cosmology*, arXiv: 2201.04741
129. D.J. Schwarz, C.J. Copi, D. Huterer, G.D. Starkman, CMB Anomalies after Planck. Class. Quant. Grav **33**, 184001 (2016). https://doi.org/10.1088/0264-9381/33/18/184001, arXiv: 1510.07929
130. C. Heymans et al., KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints. Astron. Astrophys A **646**, 140 (2021). https://doi.org/10.1051/0004-6361/202039063 arXiv: 2007.15632
131. KiDS collaboration, KiDS-1000 Cosmology: Cosmic shear constraints and comparison between two point statistics. Astron. Astrophys. A **645**, 104 (2021). https://doi.org/10.1051/0004-6361/202039070, arXiv: 2007.15633
132. DES collaboration, Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing. Phys. Rev. D **105**, 023520 (2022). https://doi.org/10.1103/PhysRevD.105.023520, arXiv: 2105.13549
133. L. Verde, T. Treu and A. G. Riess, Tensions between the Early and the Late Universe. Nat. Astron. **3**, 891 (2019). https://doi.org/10.1038/s41550-019-0902-0. arXiv: 1907.10625
134. E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri et al., In the realm of the Hubble tension—a review of solutions. Class. Quant. Grav **38**, 153001 (2021). https://doi.org/10.1088/1361-6382/ac086d arXiv: 2103.01183
135. A. G. Riess et al., A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km/s/Mpc Uncertainty from the Hubble Space Telescope and the SH0ES Team. arXiv: 2112.04510
136. N. Schöneberg, G. Franco Abellán, A. Pérez Sánchez, S.J. Witte, V. Poulin, J. Lesgourgues, The H_0 Olympics: A fair ranking of proposed models. arXiv: 2107.10291
137. T. Eifler et al., Cosmology with the Roman Space Telescope – multiprobe strategies. Mon. Not. Roy. Astron. Soc. **507**, 1746 (2021). https://doi.org/10.1093/mnras/stab1762 arXiv:2004.05271
138. L. Amendola et al., Cosmology and fundamental physics with the Euclid satellite. Living Rev. Rel. **21**, 2 (2018). https://doi.org/10.1007/s41114-017-0010-3 arXiv: 1606.00180
139. K. Nandra et al., *The Hot and Energetic Universe: A White Paper presenting the science theme motivating the Athena+ mission*. arXiv: 1306.2307

140. TMT International Science Development Teams & TMT Science Advisory Committee collaboration, Thirty Meter Telescope Detailed Science Case: 2015. Res. Astron. Astrophys **15**, 1945 (2015). https://doi.org/10.1088/1674-4527/15/12/001, arXiv:1505.01195
141. CTA Consortium collaboration, B.S. Acharya et al., Science with the Cherenkov Telescope Array. WSP, 11 (2018). https://doi.org/10.1142/10986, arXiv:1709.07997
142. J. Biteau, M. Meyer, *Gamma-ray Cosmology and Tests of Fundamental Physics*, arXiv:2202.00523
143. M. Moresco et al., Unveiling the Universe with Emerging Cosmological Probes, arXiv:2201.07241
144. P.-J. Wu, Y. Shao, S.-J. Jin, X. Zhang, *A path to precision cosmology: Synergy between four promising late-universe cosmological probes*, arXiv:2202.09726
145. D. Reitze et al., Cosmic explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO. Bull. Am. Astron. Soc. **51**, 035 (2019). arXiv:1907.04833
146. M. Maggiore et al., Science case for the Einstein telescope. JCAP **03**, 050 (2020). https://doi.org/10.1088/1475-7516/2020/03/050, arXiv:1912.02622
147. LISA collaboration, *Laser Interferometer Space Antenna*, arXiv:1702.00786
148. W.-R. Hu, Y.-L. Wu, The Taiji Program in Space for gravitational wave physics and the nature of gravity. Natl. Sci. Rev. **4**, 685 (2017). https://doi.org/10.1093/nsr/nwx116
149. TianQin collaboration, TianQin: a space-borne gravitational wave detector. Class. Quant. Grav**33**, 035010 (2016) https://doi.org/10.1088/0264-9381/33/3/035010, arXiv:1512.02076
150. LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER collaboration, A gravitational-wave standard siren measurement of the Hubble constant. Nature **551**, 85 (2017). https://doi.org/10.1038/nature24471, arXiv:1710.05835

The Search for Dark Matter, Dark Energy, Black Holes, Star Formation and Other Cosmological Searches—Accelerators

Introduction to Dark Matter Searches at CERN

Suchita Kulkarni

1 Introduction

The mystery of dark matter is related to the Universe at the largest scales and perhaps also at the smallest scales. The observable Universe, which can be described with the Standard Model (SM) of particle physics constitutes only 5% of all Universe. The rest 95% remains to be understood.

Physics gains confidence in making such profound statements by means of observations. A primary probe related to dark matter in this direction is the the Cosmic Microwave Background (CMB) spectra, which has been measured with an unprecedented accuracy [1]. The CMB was created when the Universe cooled down enough to be transparent to light and other electromagnetic radiation and it represents the furthest time and distance that can be observed in the Universe given our instruments today.[1] The angular distributions of these photons and electromagnetic radiation are affected by the constituents of the Universe. Precise measurements of these angular distributions have thus allowed us to reach the conclusion about 95% of the Universe being unknown.

Out of this 95%, about 70% is filled with so called dark energy, and the rest 25% is filled with so called dark matter. Dark matter is called dark matter because to our best knowledge it does not interact with light and does not carry SM color or electric charge.[2] An excellent example of such a particle in the SM is the neutrino. Unfor-

[1] For more on CMB physics, see the chapter "Modern Cosmology, an Amuse-Gueule" of this book.

[2] One way to get an idea of the SM color or electric charge is to think about e.g. electrons and the quarks of the Standard Model. The charge of an electron is one example of the electric charge. The SM quarks in addition also have a color charge. This one reason why the proton is made of three quarks. If the dark matter was to have any of these charges, it would interact directly with these particles and alter a lot of Standard Model physics, which is forbidden by observations.

S. Kulkarni (✉)
Institute of Physics, NAWI Graz, University of Graz, Universitätsplatz 5, A-8010 Graz, Austria
e-mail: suchita.kulkarni@uni-graz.at

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_4

tunately being very light, neutrinos are not ideal dark matter candidates. However, given that neutrino and DM are both charge and colorless particles, their signatures at the experiments can also be very similar. For example, both these particles can not be directly detected at the colliders. They escape the detector leaving behind a 'mismeasurement' in the total energy of the interaction also known as 'missing energy'.

As we know it today, dark matter is responsible for formation of the Universe via at least gravitational interactions. Being about 5 times more abundant than the ordinary matter and pressureless, dark matter created gravitational potential wells in the early Universe. Ordinary matter i.e. the SM constituents fell into these wells and formed galaxies. Today, we see this dark matter 'halo' around the galaxies when we measure the rotational curves of the stars. These rotational curves which were in-fact one of the first dark matter evidences provide a consistent picture with that inferred via CMB.

There are more dark matter evidences, ranging from the scale of galaxy to the size of the observable Universe [2]. Therefore, it is not just a singular, precise observation but a collection of (mostly) mutually consistent[3] observations that ensure us the presence of dark matter.

It is not just a singular, precise observation but a collection of (mostly) mutually consistent observations that ensure us the presence of dark matter.

One of the first dark matter evidences was pointed by Henry Poincaré and Lord Kelvin [3], who coined the term "matière obscure" and attempted to measure the amount of dark matter. This was in the year 1904 and that means that the problem of dark matter or rather the lack of its precise explanation exists in particle physics for more than a 100 years. In these 100 years, an enormous progress has been made in order to understand the dark matter properties. There are laboratory experiments such as colliders and fixed target experiments which aim at creating dark matter from SM initial states [4] and detecting the resulting missing energy signature. These are complimented by underground experiments known as direct detection experiments [5], which aim at detecting nuclear/electronic recoils of dark matter particles coming from cosmic distances and finally observations are being made at cosmic scales to detect any annihilation or decay products of dark matter. These are known as indirect detection experiments [6]. In addition to above, many of the constraints or understanding of dark matter properties come from early Universe such as the CMB or the Big Bang Nucleosynthesis (BBN) or the evolution of the large scale structure.

A few of the dark matter properties we are interested in are - what is dark matter made of, what is its origin, how and if it interacts with the Standard Model. CMB observations to date have established the exact amount of dark matter existing

[3] Currently, there are some tensions in determination of the Hubble constant using different measurements, while these may affect the exact amount of dark matter in the Universe, it will likely not eliminate the need of dark matter.

today in the Universe, also known as the relic density [7]. It is worth noting that relic density has been measured very precisely. In addition, constraints from BBN require that dark matter is mostly charge and color neutral.[4] The observations and simulations of large scale structure such as galaxies and galaxy clusters imply that dark matter travels slowly i.e. it is mostly cold. A fast travelling dark matter particle would essentially destroy the structure formation processes due to its free-streaming length. Non-observations of any direct Standard Model - dark matter interactions for dark matter searches at laboratory and cosmos imply that dark matter interacts with the Standard Model with weak interactions. Finally, there is a known discrepancy between inferred dark matter density profiles at the center of galaxies and the predictions from simulations. This is known as a "core-cusp" problem. This motivates that dark matter may have self-interactions which may alleviate this problem [8]. For a review of dark matter evidences see e.g. [2, 3, 9].

Apart from these properties we do not know the origin of dark matter or its mass. Neither do we know whether it is a single particle or made of collection of many different particles nor can we confidently say if it is a stable particle or it decays.[5] There are theoretical limits on the mass of dark matter which span some 36 orders of magnitude and their interaction cross section with SM can span some 50 orders of magnitude [10]. This vast parameter space leaves with many exciting theory ideas which need to be tested against experimental data.

We do not know the origin of dark matter or its mass. Neither do we know whether it is a single particle or made of collection of many different particles nor can we confidently say if it is a stable particle or it decays.

To date, there is no concrete evidence that dark matter is made of particles. However there are compelling reasons to believe so. One of the most interesting reasons is the conclusion that observations imply that the SM, which describes just 5% of the Universe, is incomplete. Therefore, it is not only the observation of dark matter, but also the evidence of neutrino masses, the matter-antimatter asymmetry, or the inflation, which hint at a more fundamental set of laws governing the Universe. Each of these observations is critical for the existence and evolution of the Universe and is further complemented by the Higgs hierarchy problem. The SM however does not offer an explanation for why any these phenomenon should exist. Since all these observations point to extension of the Standard Model, we enter the realm of Beyond the Standard Model (BSM) physics or simply New Physics (NP).

It is thus valid to assume that dark matter is part of a BSM theory i.e. it is made of particle(s) and understand if the particle properties of dark matter can be probed via observations. How can this obtained? The way to understand properties of such

[4] For more on BBN physics, see the chapter "Modern Cosmology, an Amuse-Gueule" of this book.

[5] We know that if the dark matter decays it has to do so with a lifetime greater than the age of the Universe. Otherwise all the dark matter in the Universe would have vanished by now.

unknown particles is via probing it's interactions with the known ones. In case of dark matter, it is done by means of probing any DM interactions with the SM particles.

If dark matter is a part of an extended Standard Model, there exists a very good possibility that dark matter interacts with the Standard Model. These interactions may even play a central role in explaining the origin of dark matter i.e. explaining the relic density. Finally, the dark matter - Standard Model interactions may also be detected at the experiments. The aim of this chapter is to concentrate on these dark matter interactions with the Standard Model particles and illustrate how and what kind of dark matter searches are being carried out at CERN.

Before moving onto introducing the experiments and what they do, it is important to illustrate what kind of dark matter scenarios are being proposed in the theoretical landscape. It is impossible to give a complete overview of such a large subject, therefore, keeping in mind the experiments covered, in the following we cover only a few examples and sketch broad aspects.

2 Particle Physics Realisations

The fact that dark matter has survived through the history of the Universe means it is either stable or has lifetime longer than the age of the Universe i.e. it lives for longer than 14 billion ($\sim 10^{10}$) years. Theorists have to hence keep the dark matter particles stable in their proposed frameworks i.e. any dark matter decays are forbidden. Mathematically speaking, this is done by means of imposing symmetries.[6] A well known example of such an externally imposed symmetry is the so called $\mathbb{Z}_2$ symmetry. Under this symmetry only pair production or annihilation of dark matter particles is possible, rendering them stable.

The expectation that dark matter - Standard Model interactions exist give rise to the experimental searches. One way to generically understand the implications of dark matter - Standard Model interactions is demonstrated in Fig. 1. In many generic dark matter theories, specially obeying $\mathbb{Z}_2$ symmetry, two dark matter particles can interact with two Standard Model particles. Let us denote such interactions by $2 \to 2$ interactions. These $2 \to 2$ interactions can destroy dark matter in the initial state to create the Standard Model particles in the Universe today. This gives rise to the indirect detection (ID) experiments. The $2 \to 2$ processes can also result in scattering of dark matter off Standard Model particles at direct detection experiments or one could hope to produce dark matter from initial state Standard Model particles at colliders. While this picture is conceptually easy to understand, concrete predictions depend on the specifics of the theory.

[6] It is not always necessary to explicitly forbid dark matter decays, in some models dark matter decays are automatically forbidden. A well known example of such occurrence in the Standard Model is for example stability of proton. We know today that proton lives for longer than 10^{34} years.

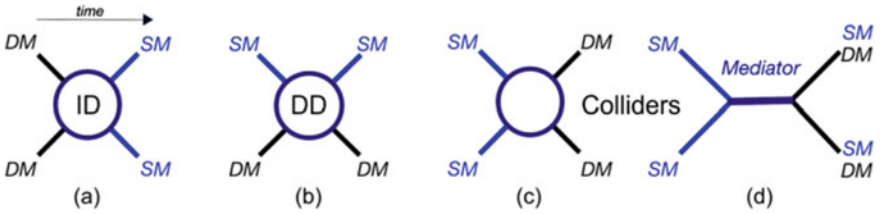

Fig. 1 Possible ways to search for dark matter across several experiments. Figure taken from [11]

The first and foremost question one could ask is how strong should these dark matter - Standard Model interactions depicted in Fig. 1 be? This question already turns out to be harder than expected. Even if one neglects the specifics of the theory, the exact strength of these interactions is not known. From experiments, it is known that this strength should be weak. One way to get a number for this strength is to demand that the same interactions are responsible for generation of relic density. As relic density is measured, the measurement fixes the interaction strength.

In technical terms, such an exercise can be done by means of solving the Boltzmann equation with dark matter - Standard Model thermal equilibrium as a starting point. Even if the details of this formula are not readily understood, it is worth looking at the equation to understand that the underlying physics is composed of many many quantities. A general expression for relic density Ωh^2 from solution of Boltzmann equation is

$$\Omega h^2 = \frac{s_0}{\rho_c}\left(\frac{45}{\pi g_*}\right)^{1/2} \frac{x_f}{M_{Pl}} \frac{1}{\langle \sigma v \rangle}, \tag{1}$$

where s_0 is the entropy of the Universe today, ρ_c is the critical density, g_* is the effective degrees of freedom, $x_f = m_{DM}/T_F$, the ratio of dark matter mass to freeze-out temperature, M_{Pl} is the Planck mass and finally $\langle \sigma v \rangle$ represents the thermally averaged dark matter - Standard Model interaction cross section. $\langle \sigma v \rangle$ includes the dark matter - Standard Model interaction strength. In this formula, Ωh^2 is known, thus if all the other numbers except $\langle \sigma v \rangle$ are known, one could predict the dark matter - Standard Model interaction strength.

Several things about this formula are to be noted. First and foremost, the dark matter relic density depends on many parameters as described in Eq. 1. Some of these parameters depend on how the Universe evolved since the very beginning, which is not precisely known yet. Second, the above arguments have implicitly assumed that dark matter - Standard Model were the only kind of processes responsible for relic density generation. This may be false if dark matter is a part of an extended sector and processes such as co-annihilation are feasible. Third, the formula assumes that dark matter - Standard Model were in thermal equilibrium, this may also not be the reality. Finally, it assumes that dark matter is made of single particle. If dark matter consists of many particles then the value of relic density per component can be just a fraction

of the observed value. It is therefore clear that one can use relic density as a guiding principle to set targets for experiments however it is useful for the experiments to search for dark matter - Standard Model interactions as generically as possible.

One can use relic density calculations as a guiding principle to set targets for experiments however it is useful for the experiments to search for dark matter - Standard Model interactions as generically as possible.

While the notion of the relic density is immensely helpful it still does not fix all the detail of the particle physics properties dark matter may have. This is controlled by the underlying theory model. Broadly speaking there are two avenues to systematically explore theory models. First, to assume that there is exactly one theory which explains all the shortcomings of the Standard Model and dark matter is part of this theory. This is the realm of 'full models'. On the other hand, it may be presumptuous to expect so much for a single theory, it is thus favourable to add only few new particles to the Standard Model and understand if such a model exists. This can be classified under the umbrella of 'simplified models'. Due to their generality, simplified models are used widely to report experimental search results.

2.1 Full Models

If the shortcomings of the Standard Model are connected to one another, then perhaps there exists a theory which can simultaneously address one or more of these problems. This theory or theories may thus also be governed by principles of symmetry which have been incredibly successful within the Standard Model so far. Such an assumption drives development of 'top-down' particle physics model developments, where multiple particles, interactions may exist. These theories have rich consequences and have an advantage of being theoretically sound. The underlying symmetries also may impose relationships between particle mass and possible decay modes. One such theory is supersymmetry (susy). The very concept that principles of symmetry may guide us to BSM physics is appealing in the bigger picture of understanding the Universe and intimately connects mathematics with particle physics at a deeper level. The model building exercises within the context of susy have revealed to us that such models may come with too many free parameters and have many possible searches at the experiments [12]. As the theories have too many free parameters, they are also difficult to asses.

2.2 Simplified Models

On the other hand, we might ask too much from nature if we assume that there exists only one theory which solves all observed mysteries. Even if such theory exists, perhaps not all new particles may be accessible to our experiments and only few new particles may be light. The theoretical prejudice that only one theory address all problems and all of its constituent particles can be discovered, may hinder the development of particle physics and may hide potential breakthrough. It thus becomes interesting to build simple models, with few new particles. Such simple models dubbed 'simplified models' are inspired by full model developments. However they do not confine themselves to strict boundaries of theoretical prejudice e.g. they may not exhibit strict relationships between particle masses. Such models because have a great advantage of being simple to understand, evaluate, cross-correlate against several experiments and free of theoretical prejudice.

On a conceptual level, building a simplified model often involves adding one mediator and one dark matter particle to the Standard Model. This may feel too simple, but often leads to very interesting consequences at colliders. To understand examples in more details see the chapter "Searching for Dark matter with the ATLAS detector" and "Searches for dark matter at CMS".

Mathematical Description of Simplified Models

For a more advanced reader it is worth spelling out some of the simplified models in mathematical language.

A very popular example of dark matter simplified model used at the LHC is extension of the Standard Model with a dark matter particle and a spin-1 gauge boson, the Z'. For these models, stability of dark matter is ensured via imposing a $\mathbb{Z}_2$ symmetry. The exact choice of the Z' coupling to the dark matter or Standard Model in the simplified model is a choice to be made. In the Standard Model sector, the Z' could couple to all fermions or just the Standard Model quarks. The LHC dark matter working group has suggested to work with the latter choice. The Z' - quark couplings could be vector or axial-vector. Therefore the relevant interaction terms in the Lagrangian is written as

$$\mathcal{L}_{\text{vector}} = -g_q \sum_q Z'_\mu \bar{q}\gamma^\mu q - g_\chi Z'_\mu \bar{\chi}\gamma^\mu \chi, \tag{2}$$

$$\mathcal{L}_{\text{axial-vector}} = -g_q \sum_q Z'_\mu \bar{q}\gamma^\mu \gamma^5 q - g_\chi Z'_\mu \bar{\chi}\gamma^\mu \gamma^5 \chi \tag{3}$$

Another very popular mediator model is the so called dark photon mediator. In this case the additional Z' gauge boson (in this case called A') has no direct couplings to the Standard Model, however generates such interactions via mixing with the Standard Model photon and Z boson. The mixing strength is denoted by ϵ. The main difference between the Z' mediator and dark photon (A') mediator model is that for Z' mediator models, one could choose to couple to only some of the Standard Model fermions, for the dark photon models described here coupling to all Standard Model

fermions is unavoidable. The corresponding gauge terms in the Lagrangian are

$$\mathcal{L}_{\text{dark}-\text{photon}} \subset -\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}A'_{\mu\nu}A'^{\mu\nu} + \frac{1}{2}\frac{\epsilon}{cos\theta}A'^{\mu\nu}B^{\mu\nu} + \frac{1}{2}m^2_{A'}A'_{\mu\nu}A'^{\mu\nu}. \quad (4)$$

In the above Lagrangian, we have not explicitly included couplings to dark matter in order to be consistent with the benchmark models used elsewhere. For an explicit case of coupling to dark matter and associated phenomenology at the LHC see e.g. [13].

Another very interesting benchmark model used at the experiments is the axion dark matter. Axions are a proposed solution to the strong CP problem in the Standard Model QCD sector. The problem itself strongly relies on the notion of symmetries. The Standard Model QCD sector allows for a CP violating term of the form

$$\mathcal{L}_{\text{strong-CP}} \subset \bar{\theta}\,\frac{\alpha_s}{8\pi}\,\tilde{G}_{\mu\nu}G_{\mu\nu}, \quad (5)$$

with $\bar{\theta}$ the CP-violating parameter, α_s the strong coupling constant and $G_{\mu\nu}$ is the Standard Model gluon field strength tensor. $\bar{\theta}$ could take any value between 0 to 2π however from neutron electric dipole moment (EDM) constraints, it appears that $\bar{\theta}$ must be very small. There is no reason for $\bar{\theta}$ to be small in the Standard Model, and hence it is known as a strong CP problem. A proposed solution to this problem is known as the axions. In this scenario, the small value of $\bar{\theta}$ is a result of breakdown of some symmetry at a very high scale. The breakdown of this new symmetry, called the Peccei - Quinn symmetry, gives rise to new particles and these new particles are known as axions. The relevant Lagrangian is

$$\mathcal{L}_a = \frac{1}{2}(\partial_\mu a)^2 - \frac{\alpha_s}{8\pi\ f_a}a\tilde{G}_{\mu\nu}G_{\mu\nu}, \quad (6)$$

where now the axion field, a dynamic parameter, has replaced the static parameter $\bar{\theta}$ and f_a represents the axion decay constant at which the symmetry is broken. The axion typically acquires mass

$$m_a = 6\,\text{eV}\,\frac{10^6\,\text{GeV}}{f_a}, \quad (7)$$

it is thus expected that axions are very light as f_a is expected to be very large.

Specific models of axions can be written down. For a review of such models see e.g. [14]. For purpose of experimental searches, it is sufficient to characterise axions to be pseudo-scalar light particles and write down a generic Lagrangian as

$$\mathcal{L}_{\text{axion}-\text{EFT}} = \frac{1}{2}(\partial_\mu a)^2 - \frac{\alpha_s}{8\pi\ f_a}a\tilde{G}_{\mu\nu}G_{\mu\nu} + \frac{1}{4}g_{a\gamma\gamma}aB^{\mu\nu}B_{\mu\nu}, \quad (8)$$

where $g_{a\gamma}$ characterizes the axion - photon coupling, couplings to Standard Model quarks may also exist however they have been neglected here. Two free parameters the axion mass controlled by f_a and the axion photon coupling $g_{a\gamma\gamma}$ will be relevant for further discussions.

3 Dark Matter at Collider: Signature Spaces

Because of the wide variety of theories, dark matter can result in many different signatures at colliders. Therefore at the LHC, dark matter is being searched for via many ways. We have already alluded to searches in the missing energy final states, in addition to this, the theories may contain long-lived particles resulting in so-called exotic final states. Broadly speaking the searches can be classified in three categories

Dark matter can lead to several interesting signatures at colliders. Searching for them constrains many theory models and in turn leads to important consequences for theoretical model building.

- **Direct dark matter production and searches via mono-X category**: Dark matter, being color and charge neutral is expected to show up at colliders by means of missing energy signature. One cannot look for events containing just the missing energy signature, instead, at colliders, searches involving some visible state in association with the missing energy are used. This visible state can be for example emission of a jet, a photon or a boson in the initial state. These searches are classified under the mono-X category.
- **Searches in prompt final states**: Along with dark matter particles, one could also search for the mediator itself in the dijet or dilepton final states by means of resonance searches. Alternatively, if dark matter is a part of an extended dark sector e.g. in case of supersymmetry, then it can also be searched for by decays of the heavier particles into dark matter.
- **Searches in exotic final states**: At colliders, the dark matter signatures further depend on the strength of the Standard Model - dark matter couplings and in general couplings between Standard Model and the dark sector particles. Let us take the simplest dark matter simplified model as described in Eq. 3. Apart from the masses of the particles involved, the Z' couplings to Standard Model and dark matter control an important quantity, namely the Z' lifetime. It depends on the masses and couplings of the particles involved. The lifetime τ_0 is inverse of total decay width Γ. Thus, longer lifetimes result from smaller decay widths. As the decay width is proportional to the Z' couplings both to dark matter and Standard Model, smaller couplings lead to longer lifetimes. Such long-lived Z' travel finite distance before decaying and lead to searches in displaced final states.

4 Looking for Dark Matter at Colliders

There are two advantages of dark matter searches at the colliders. Firstly, at colliders since dark matter is expected to be created from Standard Model initial states, one is independent of the exact amount of dark matter that exists in the Universe. Second, collider experiments offer a controlled environment for potential dark matter production. A controlled environment implies precise knowledge of the initial state particles and their energies as well as knowledge of all the known particles produced in the collision of these initial particles. This allows for more precision. Finally, different colliders use different beams at different energies, which allows to probe potential dark matter interactions across energy ranges and particle interactions. This is extremely useful given the vast dark matter parameter space.

There are two advantages of dark matter searches at the colliders. Firstly, at colliders since dark matter is expected to be created from Standard Model initial states, one is independent of the exact amount of dark matter that exists in the Universe. Second, collider experiments offer a controlled environment for potential dark matter production.

There are two primary considerations while understanding prospects of dark matter discovery at colliders. First, whether enough number of events ($\mathcal{N}$) will be produced at colliders. This can be estimated by using the formula

$$\mathcal{N} = \mathcal{L} \times \mathcal{A} \times \epsilon \times \sigma \times \mathrm{BR} \times \mathcal{P}(\mathrm{l}_1 < \mathrm{L} < \mathrm{l}_2), \tag{9}$$

where $\mathcal{L}$ is the luminosity of the collider, $\mathcal{A} \times \epsilon$ encodes the detector geometry as well as the efficiency to be able to efficiently detect final state of interest, while $\sigma \times \mathrm{BR}$ is the production cross section convoluted with the branching ratio of final state of interest. Finally, $\mathcal{P}(l_1 < L < l_2)$ denotes the probability that the a particle decays within the distance l_1, l_2. For searches in prompt final states this factor is 1, for searches in displaced final states, it depends on the particle lifetime and the region of detector where searches are carried out. From Eq. 9, it clear that having large $\sigma \times \mathrm{BR}$ is advantageous. This does not however mean that processes with small $\sigma \times \mathrm{BR}$ are out of reach. They can be probed by increasing the luminosity $\mathcal{L}$ sufficiently high. Finally, modifying detector geometry can also help gain some of the $\mathcal{A} \times \epsilon$ however this is usually tough.

An additional complication arises in the form of backgrounds. As we use SM final states to search for dark matter at colliders, any other process resulting in the same final state complicates the search. An example of such a process would be production of SM neutrinos at colliders. Similar to production of two dark matter particle which lead to missing energy at colliders, production of two neutrinos will also lead to missing energy final state. Therefore production of neutrinos is a background for dark matter final states in mono-X searches. In addition to these backgrounds, there

are also other backgrounds arising from e.g. cosmic muons or backgrounds related to detector materials. Separating signal processes from relevant backgrounds requires optimization of searches and forms one of the major efforts for collider searches.

4.1 Dark Matter Searches at Energy Frontier

The primary advantage of high energy colliders is, as the name suggests, energy. Higher beam energies allow production of heavier particles and hence allow to search for heavy dark matter or new particles.

Currently, the highest energy collider is the Large Hadron Collider (LHC), located at CERN. It collides beams of protons and operates at the center-of-mass energy $\sqrt{s}$ of 13 TeV, each beam carries the energy of 6.5 TeV and therefore the resulting collision is 'symmetric'.

The primary advantage of high energy colliders is, as the name suggests, energy. Higher beam energies allow production of heavier particles and hence allow to search for heavy dark matter or new particles.

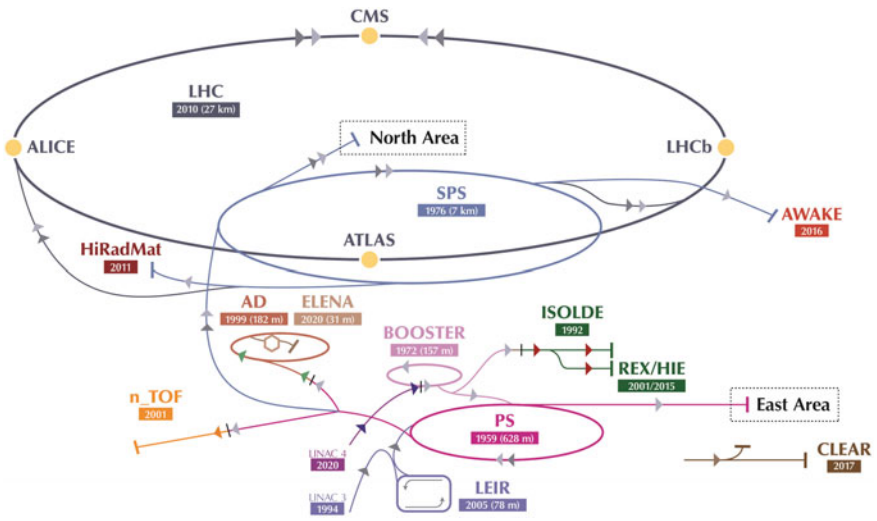

Fig. 2 A schematic view of CERN accelerator complex. Several experiments including the LHC can be seen, accelerators used to create high energy proton beams such as LINAC and PS are also seen. *Credits* CERN, figure taken from [15]

LHC houses four different experiments. There are two general purpose experiments ATLAS and CMS. The other two experiments LHCb and ALICE focus on flavour physics and heavy ion collisions and study new physics in context of these searches and measurements. The entire CERN accelerator complex including the LHC is seen in Fig. 2.

Currently, there are plans to build the next highest energy collider at 100 TeV. This will give access to unprecedented energy where even heavier dark matter can be accessed.

4.2 Dark Matter Searches at Intensity Frontier

As explained in the Eq. 9, sensitivity for low dark matter production cross sections can be obtained by increasing luminosity $\mathcal{L}$. It is usually very difficult to achieve high energy as well as luminosity at the same time.

From this point of view the intensity frontier experiments prove incredibly useful. Among these is the Belle-II experiment. This collider, situated in Japan, runs at 10 GeV center of mass energy and collides electron - positron beams. The experiment is expected to achieve an integrated luminosity of 50 ab^{-1}. For comparison the high luminosity phase of the LHC targets a luminosity of 3 ab^{-1}. The advantage of having a larger luminosity is thus clear.

Another advantage of Belle - II collider is the configuration of the initial beams. An electron - positron beam offers a much cleaner environment as compared to the proton-proton beam. This has to do with two things, first, the protons are made of three quarks and second they interact via what is known as strong interactions. As proton is made of three quarks, even if the energy of a proton in the LHC beam is precisely controlled, the energy of quarks inside the proton is not. When two protons collider, effectively the quarks inside them are responsible for final events. However as the energy of the quarks is not well controlled, every collision happens at slightly different energy. An immense theoretical and experimental effort is made to correct for this effect. Secondly, as the quarks interact via strong interactions, at the LHC a lot of jets are produced. This leads to a challenging environment at the detectors and special efforts are needed in order to navigate this. Electron - positron colliders such as Belle-II on the other hand have a much precise control over their energies and a much quieter environment as electrons do not interact via strong force.

Luminosity is a measure of how efficiently a particle accelerator produces collision events. In order to get high luminosity, more number of particles should produce collision events. This can be done by two ways, one is to create beams of highly dense particles, and another is to let the experiment run for a long period of time. The former is known as instantaneous luminosity, and the latter, integrated luminosity. It is hard to keep the experiment running

forever, therefore, creating high density beams is also preferred. It is easier to create high density beams of electrons compared to protons. On the other hand, electrons radiate more and more photons when accelerated at higher energies and thus, while electron beams can achieve higher instantaneous luminosity, achieving higher energies is more feasible with proton beams.

'Low energy' colliders such as Belle-II therefore come with significant advantages and disadvantages. On one side they offer high luminosity and cleaner environment, while on the other hand, given limited reach in the luminosity, they can probe either scenarios involving light new physics particles or theories with effective couplings.

Another kind of high intensity experiments are so-called fixed target experiments. In this setup a particle beam collides with a target stationary in the laboratory frame. In most cases, such collision produces secondary particles for specific studies. As one of the targets is stationary, fixed target experiments often lead to lower center-of-mass energy.

In order to understand this mathematically, consider a collision of two protons, with four momentum p_1^μ, p_2^μ. In one case the collision happens in the center-of-mass frame i.e. the both protons have the same energy. The four momenta can then be written as

$$p_1^\mu = (E_p, \boldsymbol{p_1}) \qquad p_2^\mu = (E_p, -\boldsymbol{p_1}), \tag{10}$$

where we assumed that both protons have the same energy. This is akin to the collider setup.

In the other case, one of the two protons is stationary, similar to the fixed target experiment. The corresponding four momenta are

$$p_1^\mu = (E_p, \boldsymbol{p_1}) \qquad p_2^\mu = (m_p, 0), \tag{11}$$

where m_p is the mass of the proton. The available energy in the case of collider system

$$s = (p_1^\mu + p_2^\mu)^2 = 2 \times E_p \tag{12}$$

The equivalent in the fixed target experiment is

$$s = (p_1^\mu + p_2^\mu)^2 = 2 \times m_p^2 + 2 \times E_p \times m_p. \tag{13}$$

Comparing the two equations, it is clear that in order to achieve the same center-of-mass energy $\sqrt{s}$, in the fixed target experiment the incoming protons will need to have a much higher energy compared to collider experiment.

However, fixed target experiments can reach rather high luminosities which can be used to probe light and/or very weakly interacting new physics particles. At CERN several fixed target experiments exist. The most notable of these for dark matter searches are NA61/62. Both these experiments use the 400 GeV proton beam from the CERN Super Proton Synchrotron (SPS) accelerator, which is then incident on the

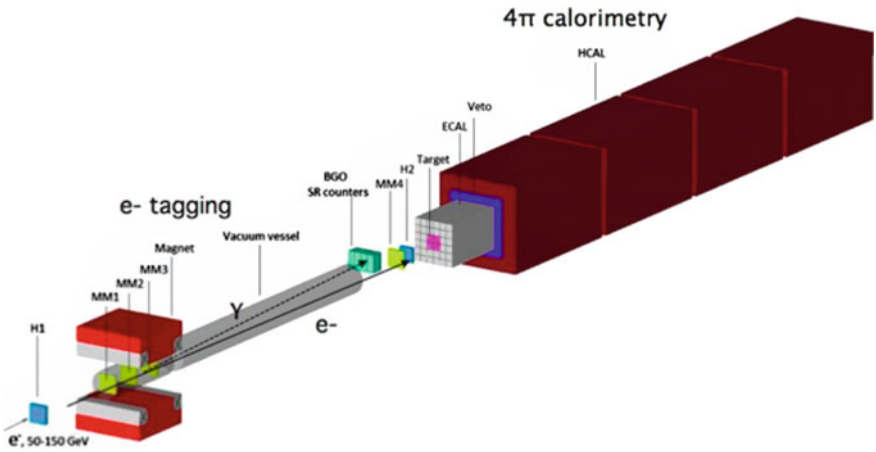

Fig. 3 Schematic view of the N64 detector. Figure taken from [16]

beryllium target. As beryllium is a composite target i.e. it has nuclei, as the proton beam collides with it, a spray of secondary particles is produced. These particles are collected downstream and analysed. Fortunately, it is still possible to make precise predictions about the composition of these secondary particles assuming Standard Model interactions. Any new physics interactions can thus show up as a deviation in these predictions and be discovered at fixed target experiments.

Another equally important experiment is the NA64 experiment. Instead of proton beam, this experiment employs the electron beam of 100–150 GeV from the Super Proton Synchrotron (SPS) machine, which is incident on a fixed target. This experiment in particular looks for light new physics particles, along with missing energy signature which can give clues for existence of dark matter (Fig. 3).

4.3 Dark Matter Searches at Lifetime Frontier

Dark matter could be a part of an extended dark sector. Should this be the case, it is equally interesting to search for particles in this extended sector. The properties of these additional dark sector particles are also unknown. While the current experimental searches demonstrate that such particles, particularly if light, should have very feeble interactions with the Standard Model sector. The exact strength of this interactions is unknown. Given the feeble interaction strength, it is likely that such additional particles are long lived and may decay outside the detector in case they are charge and color neutral. The lifetime frontier searches cover such possibilities. At the LHC, these searches and associated experiments are designed to look for very long lived particles produced in the collision of proton beams at main detectors e.g. ATLAS, CMS or LHCb. The corresponding lifetime frontier experiment situation is

some distance away from the main detectors. Any long lived charge and color neutral particle produced in proton collisions, thus escapes the main detectors and decays within the volume of the lifetime frontier experiment. Multiple proposals of lifetime frontier experiments exist. Some of them are FASER, MATHUSLA, CODEX-b, MAPP. Out of these, currently FASER has been approved and construction and first observations are ongoing. These proposals differ in their geometry, distance from main detector etc.

5 Looking for Dark Matter in Astrophysical Experiments

Astrophysical experiments aim to exploit existing dark matter in the Universe and it's potential interactions with the Standard Model in order to decode particle properties of dark matter. It is to be noted that these experiments necessarily rely on astrophysical dark matter properties such as the relic density, the exact density and velocity distribution. For some astrophysical experiments such as CAST, which will be discussed below, the details of interior of stars such as our Sun also matter.

Given the vastness of astrophysical experiments, it will not be possible to cover all of them. Instead, examples of two specific experiments, being carried out at CERN, are given here.

5.1 Alpha Magnetic Spectrometer (AMS)

Alpha Magnetic Spectrometer, better known as AMS, is an experiment aboard the International Space Station which looks for dark matter in the Universe. AMS has a dedicated control room at CERN. In particular AMS hopes to discover dark matter via observing and measuring the properties of so-called cosmic rays. Cosmic rays are high-energy particles that move through space at nearly the speed of light. Cosmic rays can be created in our own galaxy as well as in other galaxies when two Standard Model particles interact with each other in stars, active galactic nucleii, blazars etc. Cosmic rays can also be created in case dark matter produces Standard Model particles in the Universe today. One possibility for such production mechanism is illustrated in Fig. 1a. In this case $2 \rightarrow 2$ dark matter - Standard Model interactions lets two dark matter particles 'annihilate' to produce to Standard Model particles. It should be noted that dark matter decays to Standard Model final states can also produce cosmic rays.

Cosmic rays can be created in our own galaxy as well as in other galaxies when two Standard Model particles interact with each other in stars, active galactic nucleii, blazars etc. Cosmic rays can also be created in case dark matter produces Standard Model particles in the Universe today.

Whether the resulting Standard Model particles after travelling through the Universe reach the Earth or AMS is however highly uncertain. As most of Standard Model particles have charge, they are susceptible to magnetic fields in the Universe. These magnetic fields could be those in our own galaxy or extra-galactic. Due to these magnetic fields, the charged particles such as electrons, positrons get deflected. The final energy distribution as well as the total number of such charged particles arriving at the experiments is therefore dependent on our knowledge of the magnetic fields in the Universe, a task not easy to achieve as there are very few measurements of magnetic fields. The estimates of magnetic field are often performed by means of modelling and several of those exist. Some details of these models can be found in e.g. [17]. For particles which carry color e.g. the Standard Model quarks, they undergo rapid hadronization i.e. they combine in order to produce pions, protons, and nuclei. The bound states are also accelerated in the magnetic fields, they may also emit additional Standard Model particles while travelling. The details of all of this mechanism is fortunately either available or possible to obtain using experiments. The details of their propagation in the Universe, much like the charged particles, remain dependent on modelling.

While AMS measures cosmic rays sourced by Standard Model processes without dark matter initial states as well as those from dark matter, it is the dark matter sourced process which are of an interest with Standard Model sourced processes being 'backgrounds'. The details of dark matter sourced processes now depend on particle physics and astrophysics properties.

The dark matter 'source' function (Q) i.e. prediction for the number as well shape for Standard Model particles is given by

$$Q = \begin{cases} \frac{1}{2}\left(\frac{\rho}{m_{DM}}\right)^2 \sum_f \langle \sigma\, v \rangle \frac{dN^f_{e^\pm}}{dE} & \text{annihilation} \\ \left(\frac{\rho}{m_{DM}}\right) \sum_f \Gamma_f \frac{dN^f_{e^\pm}}{dE} & \text{decay,} \end{cases} \tag{14}$$

where ρ is the dark matter number density, m_{DM} the mass, and $dN^f_{e^\pm}/dE$ is the electrons or positron spectrum from dark matter annihilations in a given final state channel f. Finally, $\langle \sigma\, v \rangle$ is the dark matter interaction/annihilation cross section for the case of $2 \to 2$ processed while Γ_f is the dark matter decay width for the case of decaying dark matter.

Rather than going into the details of the above equation, it is worth noting a few characteristic things. First, the source function depends on final state being observed, the above formula is for electron/positron final states, but similar functions can also be written down for other final states e.g. proton/anti-protons. Second, the source function depends on whether the dark matter annihilates i.e. undergoes $2 \rightarrow 2$ interactions or decays. The function furthermore depends on the exact amount and the behaviour of dark matter density distribution ρ and the dark matter mass m_{DM}. The density distribution ρ is an astrophysical input and also carries uncertainty as it has not been measured very well e.g. at the center of galaxies as of yet. Finally, as the source function depends on the dark matter density, it makes sense to target regions where dark matter density is very high e.g. dwarf galaxies or at the center of our own Milky Way. This is often one of the targets of experiments such as AMS.

The AMS experiment aims to measure cosmic rays and can separate them based on their mass and charge. The experiment collects cosmic rays from all sources. This means it observes cosmic rays within our galaxy as well as from outside our galaxy. This is an incredible wealth of information, which is being used by dark matter theorists all over the world to test potential signatures of dark matter particles.

5.2 CERN Axion Solar Telescope (CAST)

The CERN Axion Solar Telescope (CAST) is dedicated to search for axions, which were described in Sect. 2.2. Compared to other dark matter candidates discussed before, axions are extremely light. As a rough comparison, the candidates being searched for at the colliders or via the AMS experiments are somewhere in GeV to $O(100)$'s GeV range, the axions on the other hand could be as light as few eV. In addition, they may interact with the Standard Model particles very weakly. A different search strategy for such light particles is therefore necessary.

The CAST telescope is one such experiment that relies on the fact that axions generically couple to photons as shown in Eq. 8. The axion photon coupling gives rise to what is known as the Primakoff effect. Primakoff effect is essentially a conversion of an axion to a photon in a magnetic field, or vice versa. This is possible because photon is particle associated with electromagnetic forces. Thus one could think about photon in terms of particles or in terms of underlying electric and magnetic fields.

The Primakoff effect can thus be used to convert a source of photons into axions and back. The photon source could for example be a powerful laser beam in the laboratory. This has the advantage of a controlled environment where one knows the energy of photon source, the exact configuration of magnetic fields etc. However these will be only as powerful as the source and the magnetic fields are.

The CAST telescope is an experiment that relies on the fact that axions generically couple to photons, which gives rise to the so-called Primakoff effect.

Another complementary approach is to use one of the stars. The star itself acts as a powerful photon source and its magnetic field can convert the photons to axions. These axions then travel to Earth where a magnetic field inside telescope can convert the axions back to photons.

The CAST telescope operates exactly on this principle. It uses the Sun as a source of photons and the solar magnetic field as a way to potentially convert the solar photons into axions. Such a principle is advantageous due to abundant solar photons and large magnetic fields. The disadvantage being, the number of axions produced in the Sun now depend on the solar photon spectrum and knowledge of solar magnetic fields.

Despite these difficulties the axion searches via Solar observations, also known as helioscope searches, are incredibly useful as they probe a large range of very light axion masses. The final number of photons observed by the telescope are given by

$$N_\gamma = \int \frac{d\phi_a}{dE} P_{a\to\gamma}\, A\, t\, dE \tag{15}$$

where, $d\phi_a/dE$ is the axion spectrum as expected at the Earth, $P_{a\to\gamma}$ is the probability of the axion-to-photon conversion, t the observation time and A the axion sensitive area of the magnet aperture. It is therefore clear that larger telescopes with long observation times are beneficial.

If axion dark matter exists, then it can be everywhere. Can one use CAST to look for such axions? Recently, the telescope was transformed into an axion haloscope for this purpose. This is done through microwave cavities which were installed for the first time inside a dipole magnet such as CAST. This technique still uses the axion - photon conversion principle explained above but the conversion is further enhanced inside a microwave cavity (such as CAST-CAPP) resonating at the axion mass.

In addition to all the above experiments, there is also the 'Baryon anti-baryon symmetry experiment' or in short BASE experiment. This experiment primarily searches explanation of why there is more matter in the Universe than antimatter. It does so by precisely measuring properties of anti-protons. In particular, BASE is equipped to trap anti-protons, which can also be used for other purposes. As an advantage of this anti-proton trap, the experiment also searches for dark matter particles which may preferentially couple more to anti-protons. One such example is axions, which may couple to anti-protons and hence can be searched for at BASE.

6 Conclusion

Despite being a century old problem, we have so far not found a convincing signal of dark matter at any of the experiments. Despite this, we have learned a great deal about dark matter by means of these experiments, astrophysical observations and combination of them. These efforts assume that dark matter is made of particles,

while there is no concrete evidence for this, there is very good motivation for such an expectation.

Within our current best knowledge, we understand that dark matter does not interact with light, we know that it is either stable or lives longer than the age of the universe. We also know that it interacts with the Standard Model particles very weakly.

This knowledge however still leaves a great deal of possibilities for dark matter candidates. These possibilities in turn give rise to many experimental searches. Many of these are being carried out at CERN, either at unprecedented energies or luminosities or via astrophysical observations.

In the end, the question remains, what do we intend to learn out of these experiments? The obvious answer is of course that we find what dark matter is. However even in the absence of signal these experiments teach us a great deal about what dark matter can not be. This knowledge is valuable as it drives further developments of theory as well as new experiments.

It is certainly true that the search for dark matter matter is far from over. However the coming years will bring an exciting era in terms of ever higher energy and higher luminosity experiments. These will be complemented by astrophysical observations. While it is impossible to guess the results of the experiments, the new knowledge gained will certainly shed more light on the properties of dark matter.

References

1. Y. Akrami et al., Planck 2018 results. I. Overview and the cosmological legacy of Planck. 7 (2018)
2. G. Bertone, D. Hooper, J. Silk, Particle dark matter: evidence, candidates and constraints. Phys. Rept. **405**, 279–390 (2005). https://doi.org/10.1016/j.physrep.2004.08.031
3. G. Bertone, D. Hooper, History of dark matter. Rev. Mod. Phys. **90**(4), 045002 (2018). https://doi.org/10.1103/RevModPhys.90.045002
4. A. Boveia, C. Doglioni, Dark matter searches at colliders. Ann. Rev. Nucl. Part. Sci. **68**, 429–459 (2018). https://doi.org/10.1146/annurev-nucl-101917-021008
5. M. Schumann, Direct detection of WIMP dark matter: concepts and status. J. Phys. G **46**(10), 103003 (2019). https://doi.org/10.1088/1361-6471/ab2ea5
6. J.M. Gaskins, A review of indirect searches for particle dark matter. Contemp. Phys. **57**(4), 496–525 (2016). https://doi.org/10.1080/00107514.2016.1175160
7. N. Aghanim et al., Planck 2018 results. VI. Cosmological parameters. 7 (2018)
8. M. Kaplinghat, S. Tulin, Yu. Hai-Bo, Dark matter halos as particle colliders: unified solution to small-scale structure puzzles from dwarfs to clusters. Phys. Rev. Lett. **116**(4), 041302 (2016). https://doi.org/10.1103/PhysRevLett.116.041302
9. G. Bertone, T.M.P. Tait, A new era in the search for dark matter. Nature **562**(7725), 51–56 (2018). https://doi.org/10.1038/s41586-018-0542-z
10. H. Baer, X. Tata, *Dark matter and the LHC*, pp. 179–203 (2009). https://doi.org/10.1007/978-81-8489-295-6_12
11. Possible dark matter signature spaces. https://atlas.cern/updates/feature/dark-matter
12. M. Tanabashi, K. Hagiwara et al., Review of particle physics. Phys. Rev. D **98**, 030001 (2018). https://doi.org/10.1103/PhysRevD.98.030001

13. J.A. Evans, S. Gori, J. Shelton, Looking for the WIMP next door. JHEP **02**, 100 (2018). https://doi.org/10.1007/JHEP02(2018)100
14. L. Di Luzio, M. Giannotti, E. Nardi, L. Visinelli, The landscape of QCD axion models. Phys. Rept. **870**, 1–117 (2020). https://doi.org/10.1016/j.physrep.2020.06.002
15. R. Vanden Broeck, The cern accelerator complex. Complexe des accélérateurs du CERN. Sep (2019). https://cds.cern.ch/record/2693837
16. Schematic view of NA64 detector. https://na64.web.cern.ch/node/8
17. P. Salati, Indirect and direct dark matter detection. PoS, CARGESE2007: 009 (2007). https://doi.org/10.22323/1.049.0009

Searching for Dark Matter with the ATLAS Detector

Caterina Doglioni and Dan Tovey

1 Introduction

Experiments [1] at particle accelerators have revealed much about the nature of visible (ordinary) matter, starting from the first prototypes that aided the discovery of the proton and the antiproton [2] to the recent discovery of the Higgs boson [3, 4]. All of the particles observed so far are part of the Standard Model of Particle Physics, describing the fundamental components of matter and their non-gravitational interactions.

As discussed in the Introduction chapter, the most powerful accelerator ever built is the Large Hadron Collider (LHC) at CERN in Geneva [5, 6], accelerating protons and colliding them with a total energy of 13 TeV. According to Einstein's most famous equation, $E = mc^2$, the more energy (E) the more massive particles (with a mass m) one can create (13 TeV corresponds to roughly 14 thousand times the rest mass of a proton). The hope is that at the LHC we can create massive dark matter particles by colliding known particles, in the same way we create the Higgs boson in proton-proton collisions.

Particles are regularly accelerated to very high energies in the universe in "natural" particle accelerators, such as supernovae explosions, and then collide with other particles in our atmosphere. Cosmic rays, for example, are particles that are generated in outer space and make it to Earth. However, the advantage of laboratory particle accelerators such as the LHC is that there we know the initial conditions of the collisions - namely the type and energy of the particles being collided. We can

C. Doglioni (✉)
University of Manchester, Manchester, UK
e-mail: caterina.doglioni@cern.ch

Lund University, Lund, Sweden

D. Tovey
University of Sheffield, Sheffield, UK
e-mail: dan.tovey@cern.ch

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_5

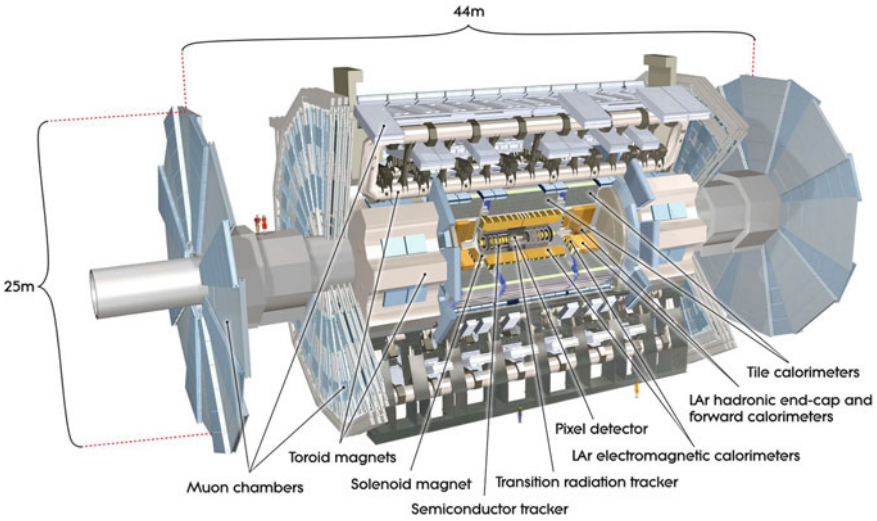

Fig. 1 Computer generated image of the whole ATLAS detector [9]. ©CERN

also create a large (and known) number of collisions and observe them in a controlled environment. These are essential features for detecting dark matter particles at collider experiments like ATLAS (Fig. 1).

With experiments at the Large Hadron Collider like ATLAS, we hope to further understand the nature of dark matter by producing and observing it in controlled laboratory conditions.

The ATLAS experiment depicted in Fig. 2 [7, 8] is located at one of the collision points of the Large Hadron Collider, and it is one of the largest particle detectors ever built.

The layout and design of the ATLAS experiment allow the detection of the LHC collision products, in the form of particles with sizable interactions with the detector material. Scientists use ATLAS data to study a wide variety of physics processes.

2 How Dark Matter Appears in the ATLAS Detector

Since dark matter is *dark*, it will not interact significantly with instruments made of ordinary matter. For this reason, the underlying signature of dark matter production at the LHC, used by ATLAS searches, is the presence of new invisible particles in proton-proton collisions. One might reasonably ask how invisible particles can be

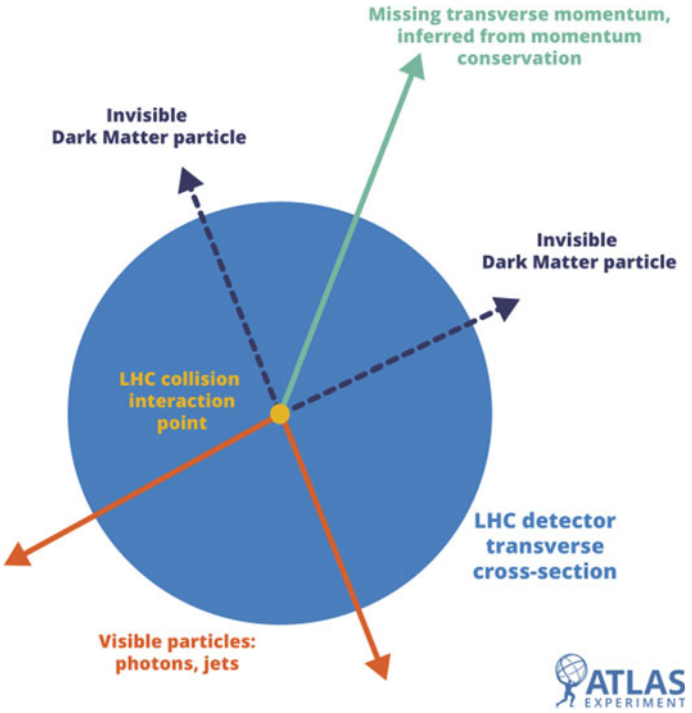

Fig. 2 Diagram showing how missing transverse momentum (ETmiss) is determined in the transverse cross-section of a LHC detector. The LHC beams are entering/exiting through the plane. (Image: C. Doglioni, L.T. Wang and E. Ward/ATLAS Collaboration, ©CERN)

observed, since they are by definition undetectable! We solve this problem with a little ingenuity.

Before each collision, the protons travel along the direction of the LHC beams, and not in directions perpendicular to the beams. This means that their momenta in these perpendicular directions – their "transverse momentum" – is zero. A fundamental principle of physics is that momentum is conserved and so, after the collision, the sum of the transverse momenta of the products of the collision should still be zero. Therefore, if we add up the transverse momenta of all the visible particles produced in the collision and find it not to be zero, then this could be because we have missed the momentum carried away by invisible particles. This happens routinely in ATLAS, in the case of physics processes involving neutrinos. We refer to this missed transverse momentum as "*ETmiss*" or *MET*.

LHC searches for dark matter look for collisions with large values of ETmiss, where the dark matter is produced in association with other visible particles from the Standard Model, such as photons, quarks or gluons (forming "jets" of particles), or electrons, muons or tau leptons. While ETmiss can be difficult to measure because

it relies on accurate measurements of all the other particles in the collision, it is a powerful tool for observing otherwise invisible particles such as dark matter.

LHC searches for dark matter look for collisions with large values of ETmiss, where the dark matter is produced in association with other visible particles from the Standard Model

A further requirement for the identification of dark matter particles in collisions is that the invisible particles should not decay as they travel through the ATLAS detector. In order for an invisible particle to be a candidate of the "relic" dark matter produced in the Big Bang, it should have a lifetime of at least the age of the Universe - of the order of 1 billion years. Particles created in LHC collisions take about 40 nanoseconds to cross the ATLAS detector, so requiring that their lifetime be longer than this is not enough, on its own, to prove they constitute the dark matter. Complementary information from astroparticle experiments searching for relic dark matter is required, as described in Sect. 5. However, hints of an invisible particle detected at one of the LHC experiments would provide very good hints for more targeted searches.

It is worth noting that other particles that are connected to dark matter might also be detected at the LHC, for example new short-lived particles that can decay both into dark matter and into known matter, or new particles that are predicted to be present together with the dark matter ones. Observing those would be an important complement to an observation of dark matter particles from space, as it would allow us to better understand the full landscape of dark matter interactions.

3 What Kinds of Dark Matter Could ATLAS Detect?

Experimentally, there are very few indications of what dark matter might be. We can, however, make theoretical hypotheses on the nature of dark matter, which are useful to experimentalists, as discussed in detail in Chap. 4. The theorist and experimentalist communities often collaborate. For example, the community of theorists and experimentalists looking for dark matter at the LHC has joined forces, forming first the Dark Matter Forum and then the Dark Matter Working Group. The goal and results of those group are described in Ref. [10].

Theoretical models of dark matter can tell us more about how the interaction of dark matter with ordinary matter may take place. From that, we can predict what to expect in our detectors if that model were realised in nature.

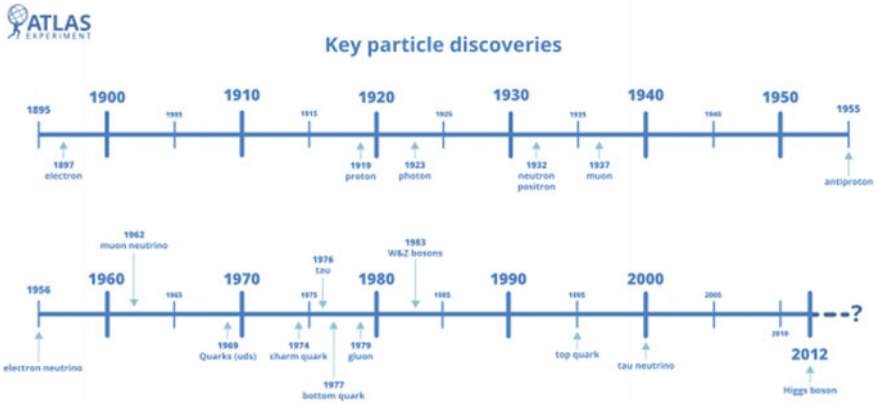

Fig. 3 A timeline showing the key discoveries of known particles (Image: E. Ward/ATLAS Collaboration, ©CERN)

Theory predictions for dark matter models is relevant for detector design, and for deciding how to record and analyse the products of the collisions. It is generally useful to know what to look for in advance, as we would not be able to collect LHC data in its entirety and must decide in real-time which collision events to record - this is done using the ATLAS trigger system. A solid theoretical framework is also necessary to put LHC searches into context, and compare them with dark matter searches from other instruments.

Searches for dark matter at the LHC are commonly guided by theoretical models that would allow us to explain the relic density of dark matter with one or a few kinds of particles. A class of models that satisfies these requirements includes a dark matter particle that only interacts weakly with ordinary particles and has a mass within the energy range that can be probed at the LHC - a Weakly Interacting Massive Particle (WIMP). Using WIMP models as our starting point for LHC searches doesn't mean that we are bound to the idea that dark matter should be described with a single particle and a single interaction. This is especially important when one considers that the content of dark matter in the Universe is five times the content of ordinary matter, and ordinary matter is described by a variety of different particles and interactions.

At the LHC, we are looking for a variety of possible theoretical models of dark matter hoping that the few most prominent components and interactions of dark matter will be detected first, just as the electron, proton and electromagnetic interaction were discovered before all other particles of the Standard Model, as shown in Fig. 3.

The simplest models one can build in terms of particle content are those where only a new kind of dark matter particle is added to the Standard Model. In these models, the interaction between visible and dark matter must proceed through existing particles, such as the Z or Higgs boson. This means that the Z or Higgs boson could decay

into two dark matter particles,[1] in addition to their ordinary decay modes involving Standard Model particles.

These models are called "portal" models of dark matter, as known particles act as the portal between what we know (ordinary matter) and what we don't know (dark matter). While models with a Z boson portal are fairly constrained by precision measurements, including those done at the LEP collider at CERN during the 1990s, now is the first time in the history of particles that we can study the properties of the Higgs boson in detail. We could discover whether one or more of those properties leads to a connection to dark matter. Searches for dark matter interacting with Standard Model particles through the Higgs portal are described in Sect. 4.1.

In addition to dark matter, one can also conceive another particle not included in the Standard Model that acts as a portal particle. These are also called "portal" or "mediator" particles, since they mediate a new interaction between ordinary matter and dark matter. In the simplest versions of these models, the mediator is an unstable heavy particle that is produced directly from the interaction of Standard Model particles, such as quarks at the LHC or electrons in a secondary interaction [11]. Therefore, the mediator must also be able to decay into those same particles that were responsible for its production, in addition to into a pair of dark matter particles (see e.g. Refs. [12–14]). If a model of this kind occurs in nature, we have a chance to directly discover this mediator particle at the LHC, as we would be able to detect its Standard Model decay products. This is only an example of a *simplified model* that is commonly used to interpret the results of many LHC searches in terms of dark matter, even though it is often deemed too simple to represent the full complexity of a dark matter theory. However, simplified models are still useful as building blocks for more complete theories with more ingredients. Examples of ATLAS searches for dark matter interacting with Standard Model particles through a new mediator particle are described in Sects. 4.2 and 4.3.

The most popular example of a more complete theory that includes a dark matter candidate is supersymmetry (SUSY). SUSY was one of the first dark matter models to be studied extensively at the LHC. An appealing feature of supersymmetry is that it also solves a stability problem of the relatively low mass of the Higgs boson and other electroweak particles of the Standard Model (around 100 GeV) compared to the Planck scale (1019 GeV), at which gravity is expected to become strong and the Standard Model must break down. Quantum field theories like the Standard Model naturally prevent such large differences in energy scale from developing, so a physical mechanism is required to generate them. SUSY models provide such a mechanism and, in many cases, predict the existence of a new stable, invisible particle - the lightest supersymmetric particle (LSP) - which has exactly the right properties to be a WIMP dark matter particle. The search for particles predicted by SUSY is a major focus of the ATLAS physics programme. If produced in LHC collisions, these particles could decay to produce a variety of Standard Model particles that can be observed in the ATLAS detector, together with two escaping LSP dark matter particles that generate

[1] This can happen if the dark matter particle mass is less than half of that of the Z or the Higgs boson.

the characteristic ETmiss signature discussed above. Examples of ATLAS searches for supersymmetric dark matter are described in Sect. 4.4.

Many other theories, of various degrees of completeness and complexity, contain dark matter particle candidates. Some of them predict new particles similar to the Higgs boson that can decay into dark matter, while others go beyond the WIMP paradigm and include mediators with extremely feeble interactions with known particles that only decay after travelling significant distances inside (or outside!) the detector, or more complex sectors of particles mirroring the Standard Model. It is important for LHC searches to cover all this ground, while also preparing for unexpected, not-yet-theorised discoveries. No stone must be left unturned, and some example searches for dark matter particles beyond the WIMP and SUSY paradigms are described in Sect. 4.5.

4 ATLAS Experimental Techniques and Results

4.1 Searches for Dark Matter Particles Associated to the Higgs Boson

Since the addition of the Higgs boson to the Standard Model after its discovery in 2012, an interesting question to ask is whether the Higgs boson has a privileged role in propagating the interactions between ordinary and dark matter. This would make LHC experiments an ideal ground to test such interactions, as the study of Higgs properties and decays is one of the physics priorities for the next decade. In the simplest scenario, the Higgs boson could be the only particle in the Standard Model that is able to decay into dark matter, leading to a minimal extension of the Standard Model that only requires the addition to a dark matter particle. More complex scenarios involve multiple Higgs-like bosons with a variety of interactions with both dark and ordinary matter in addition to the Standard Model one (see e.g. [15, 16]).

> Since the discovery of the Higgs boson, we seek whether the Higgs boson has a privileged role in propagating the interactions between ordinary and dark matter.

We have two handles to understand whether the Higgs boson decays into dark matter particles. Firstly, decays of the Higgs boson into invisible particles in the Standard Model (neutrinos) are extremely rare. Therefore, if Higgs decays to invisible particles are observed, for example via an excess of missing transverse momentum in events identified as containing a Higgs boson, they could be attributed to dark

matter.[2] Secondly, Higgs decays to dark matter particles would change the rates at which the Higgs boson decays into other particles, since the total probability of the Higgs decaying into any other particle must be equal to unity. The most recent ATLAS results to date have not observed any signs of extra invisible decays of the Higgs boson in either of these two approaches, and can place strong constraints on dark matter particles interacting exclusively with the Higgs boson [17].

4.2 From Measuring Invisible Particles to MET+X Searches

ATLAS measures many processes involving invisible Standard Model particles: neutrinos. Figure 4 shows the results of the measurement of the number of Z bosons decaying into a pair of neutrinos (about one fifth of all Z boson decays).

As shown in the diagram in Fig. 5, we use a particle that interacts with the detector and is therefore visible (in this case a photon) to detect the presence of invisible particles and measure their missing transverse energy, as explained in the previous section.

A very similar technique can be used for detecting the presence of dark matter particles. If we take the process in Fig. 5, replace the neutrinos with dark matter particles, replace the Z boson with a generic mediator between ordinary matter and dark matter, then we have the diagram in Fig. 6.

What appears in the ATLAS detector (also called detector *signature*) for the two processes shown in Figs. 5 and 6, and it is shown in Fig. 8: energy deposited by the photon and nothing else, leading to MET back-to-back with the photon (Fig. 7).

A variety of techniques, including machine learning algorithms, are used to precisely measure the properties of the particles contained in the collision events (including the MET). Since we cannot distinguish the processes on a collision-by-collision basis, we have to take a different approach. We start by collecting a large number of events that have a large amount of missing transverse momentum and a highly energetic particle (or group of particles). Then, we estimate precisely the number of expected events from known Standard Model processes (the *backgrounds*), and look for an excess of additional events that could be due to dark matter processes. This kind of search is called a "MET+X" search, where MET stands for missing transverse momentum and X stands for what the dark matter recoils against.

This kind of MET+X search makes no specific assumption about the nature of the invisible particles, other than that they are produced in association with a Standard Model particle. It is therefore well-suited to cast a wide net on a variety of dark matter models, as long as the model's signature includes invisible particles and includes dark matter-Standard Model interactions. Conversely, the very large Standard Model

[2] One of the physics targets for the high luminosity LHC era and future colliders is still to measure the rate of the decays of the Higgs boson into neutrinos and compare it to theoretical predictions, but since those decays make up less than a thousandth of the total possible Higgs decays, we will need to wait until enough Higgs bosons are collected!

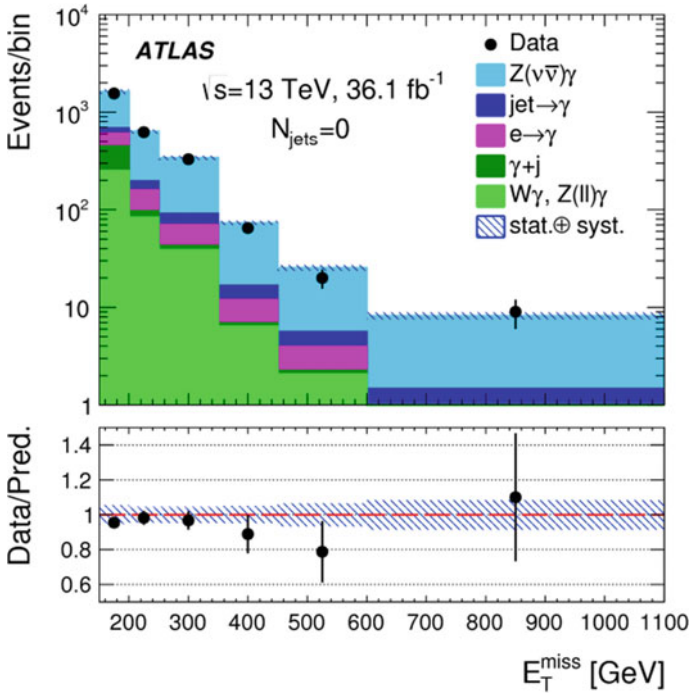

Fig. 4 Measurement of the number of events in data for each range of missing transverse momentum, compared to the sum of different physics processes that produce this signature in the detector, from simulation. Figure from [18] and ©CERN

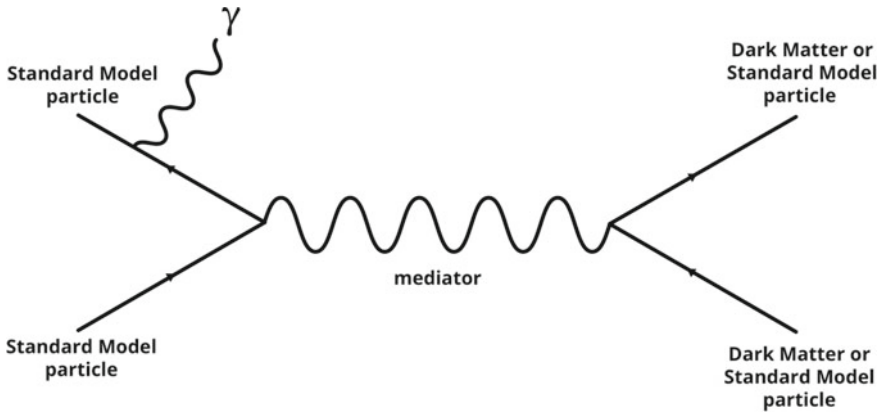

Fig. 5 Diagram of a Z boson decaying into a neutrino-antineutrino pair where the Z boson is produced in association with a photon. ©CERN

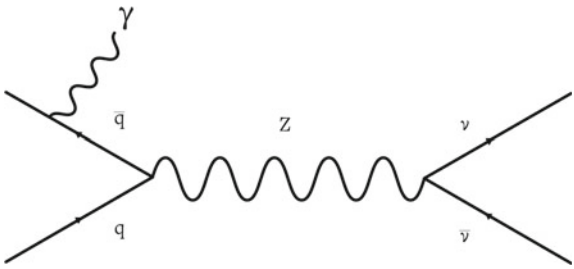

Fig. 6 Diagram of a new mediator particle decaying into a pair of dark matter particles, produced in association with a photon. ©CERN

backgrounds in MET+X searches can be reduced by giving up some of their generality, for example by requiring distinctive particles (e.g. top quarks, the Higgs boson or related particles) to be produced in association with the dark matter.

MET+X searches make no specific assumption about the nature of the invisible particles, other than that they are produced in association with a Standard Model particle. They are therefore well-suited to cast a wide net on a variety of dark matter models.

ATLAS has searched for excesses of invisible particles recoiling against photons [19], jets [20, 21], both photons and jets [22], Z and W bosons [17, 23] as well as in association with heavy flavour quarks [24–26].

So far, we have not found any excess with respect to backgrounds in any of these searches, as shown in Fig. 9 for the MET+photon search, where the data agrees with the Standard Model-only prediction. Still, the journey of MET+X searches at the LHC is far from over. Adding data and improving the experimental precision of future searches will enable us to search for even weaker dark matter interactions yielding processes that are even rarer than the processes we are sensitive to right now.

4.3 From MET+X Searches to Searches for Visible Mediator Decays

The mediator in Fig. 6 can also decay to visible particles, leading to a peak or "resonance" in the total (invariant) mass of those particles, as in Fig. 9.

Searches for new particles using resonances in the total mass of visible particles have led to numerous discoveries at colliders, including, most recently, the Higgs boson at the LHC. Given that the LHC is the highest-energy laboratory particle collider, the most obvious goal is to search for extremely massive particles that could not have been produced before.

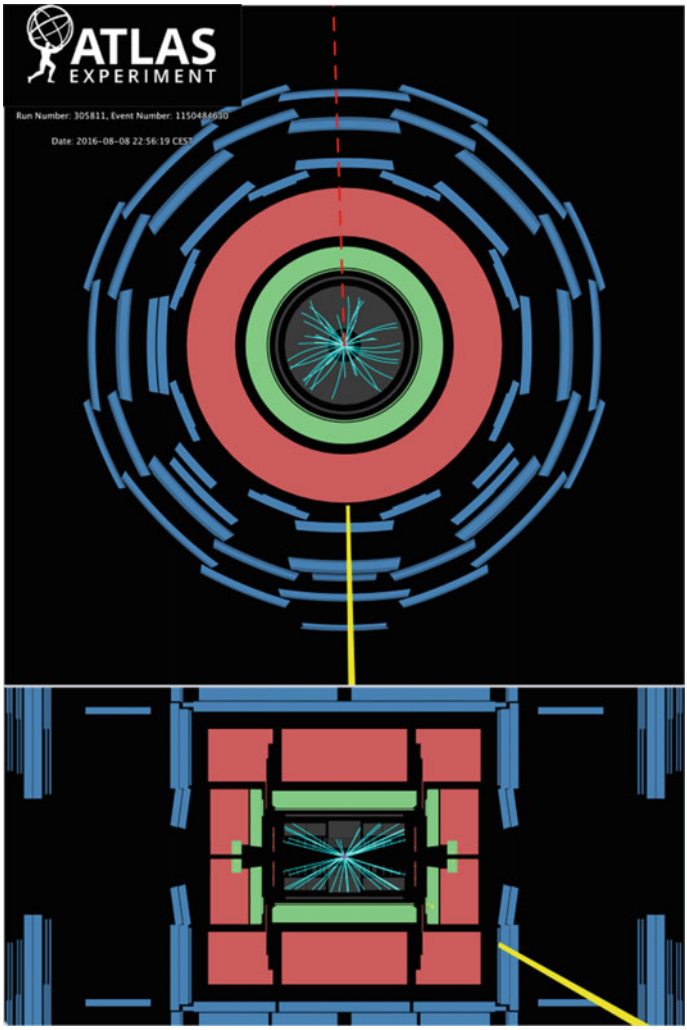

Fig. 7 A visualisation of a photon + ETMiss event recorded in 2016 (event 1150484630, run 3058110), is shown in two different sections of the ATLAS detector. A photon with transverse momentum of 265 GeV (yellow bar) is balanced by a ETmiss of 268 GeV (red dashed line in the opposite side of the detector). (Image: ATLAS Collaboration/CERN, ©CERN). Reference should be taken from https://inspirehep.net/literature/1591328

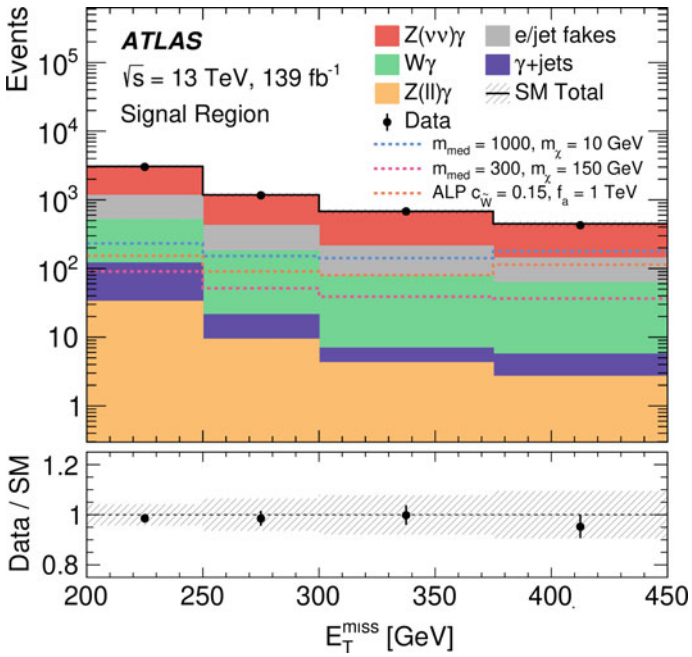

Fig. 8 Missing transverse momentum distribution in data after selecting events with an energetic photon and ETmiss, compared to the Standard Model predictions. The different background processes are shown in different colours. The expected spectra of an example WIMP dark matter scenario is illustrated with blue and red dashed lines. Figure from [19], ©CERN

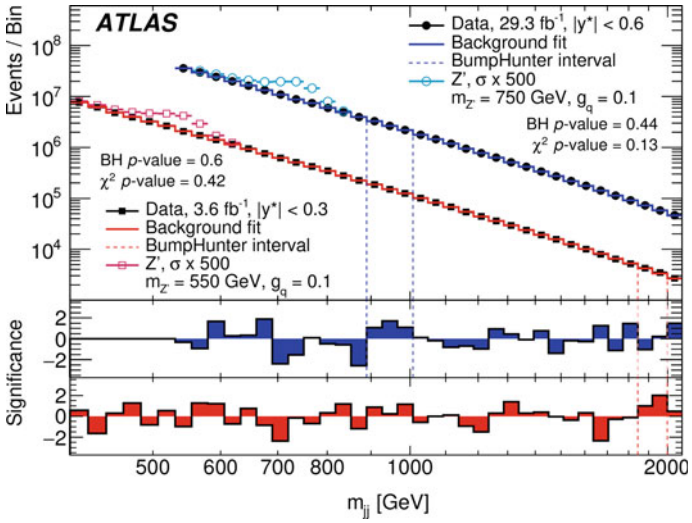

Fig. 9 Distribution of the invariant mass of two jets in ATLAS data (filled circles) with two different data analysis selections, with simulated resonance peaks for dark matter mediators overlaid in blue and red open circles lines. Figure from [27], ©CERN

Still, dark matter mediators could also appear at lower masses, escaping our eyes because of very low couplings to the protons' constituents. This is a region where it has been increasingly difficult to perform searches due to the overwhelming backgrounds that exceed the experiment's data capacity if recorded in their entirety. Since background events are indistinguishable from events coming from decays of dark matter mediators, there is a risk of discarding both. Being able to detect this kind of process has provided motivation for overcoming technical limitations. All the main LHC experiment now employ data-taking techniques that allow them to retain a smaller amount of information from the collision than in ordinary data taking, so that more events can be recorded (see Refs. [27, 28] for ATLAS). These searches have not yet yielded any new particles as one can see in Fig. 9, but improvements to the data acquisition and trigger systems may bring surprises for the next LHC run.

The state of the art of MET+X searches and searches for visible mediator decays in ATLAS as of 2018 was summarized in Ref. [29].

The results of searches for invisible and visible dark matter mediator decays bring complementary information on different parameters of dark matter models. Together, they could help to characterise the nature of a discovery.

The results of searches for invisible and visible dark matter mediator decays bring complementary information on different parameters of dark matter models. Together, they could help to characterise the nature of a discovery.

We must keep in mind, though, that the interpretation of those searches is in terms of the processes in Fig. 6, which stems from a very simple theoretical model. In this model, the only two new particles are the dark matter and the mediator of the interaction, and that may not describe the full complexity of the unknown matter in the universe.

This is why ATLAS searches target many other experimental signatures in addition to MET-X and resonance searches. For example, models including putative new Higgs bosons yield an assortment of detector signals that can be targeted by different searches. All experimental results targeting the same model can be compared to see whether there are regions in the model parameter space where we haven't yet looked and, in some cases, they can be combined to strengthen the discovery potential or constraints on dark matter models.

4.4 Searches for Supersymmetric Dark Matter Particles

Compared with the MET+X searches described above, detector signatures from SUSY scenarios offer the possibility to make use of some additional tricks to identify a potential dark matter signal from the Standard Model background. In many models, SUSY particles are produced in pairs due to a requirement to conserve a quantity

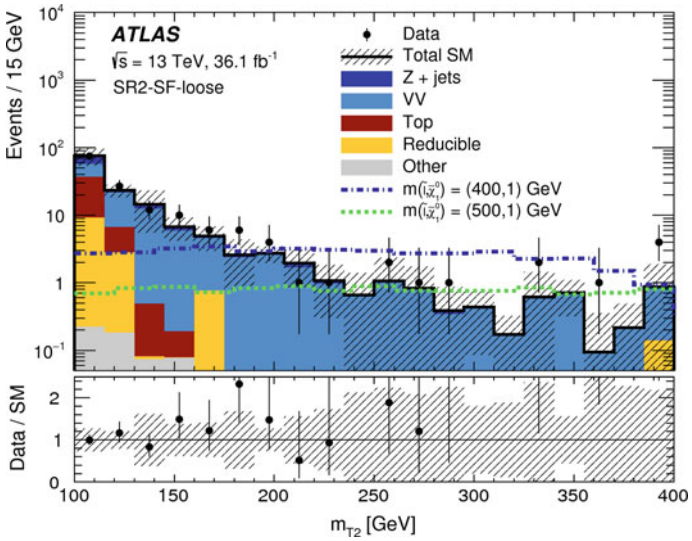

Fig. 10 Distribution in data of a quantity sensitive to the production of pairs of SUSY particles whose decays include dark matter particles, after selecting events with two electrons or muons and ETmiss, compared to the Standard Model predictions. The different background processes are shown in different colours. The expected spectra of example SUSY dark matter scenarios are illustrated with blue and green dashed lines. Figure from [31], ©CERN

called "R-parity" (sometimes also denoted "matter-parity").[3] Whenever a SUSY particle decays, the resulting decay products must include exactly one lighter SUSY particle. The decay chain ends when one of the lightest SUSY particles, which is a candidate dark matter particle, is produced. In contrast to many non-SUSY dark matter models, SUSY particle decays can generate many visible Standard Model particles of high energy. Hence events containing SUSY particles can be identified by requiring these particles as well as missing transverse momentum. A further trick is to make use of constraints on the momenta of the visible particles produced in the SUSY decays coming from the high masses of their SUSY particle parents. In particular, when two visible particles are produced from two identical decay chains in a SUSY event, we can measure properties of the event which can take on much larger values than those expected in Standard Model background events. An example result of a SUSY dark matter search after all analysis selections is shown in Fig. 10 and described in [30, 31].

With the help of these tools, SUSY searches are able to set tight requirements for events with a given set of characteristics, targeting specific models. This makes them less general than MET+X searches, but also less impacted by large numbers of background events.

[3] R-parity ensures that in SUSY models protons, and hence all of the atoms in the universe, are unable to decay to other particles quickly by exchanging SUSY particles. In models without R-parity conservation, this can also be prevented. However, introducing R-Parity is the simplest possibility.

ATLAS has not yet found evidence of SUSY LSPs, and has strongly constrained many of the models that would simultaneously solve the dark matter puzzle and provide an explanation for the low mass of the Higgs boson. Nevertheless, many SUSY variants remain interesting and the search isn't over, as described in Ref. [32]; other recent results are mentioned in Ref. [33].

ATLAS has not yet found evidence of SUSY LSPs, and has strongly constrained many of the models that would simultaneously solve the dark matter puzzle and provide an explanation for the low mass of the Higgs boson.

4.5 Searches for Non-WIMP Dark Matter Particles

Many other searches for particles from more complex dark matter theories are also performed in ATLAS even though we don't cover them in detail in this chapter. Some of the characteristics of these particles make them behave very differently compared with the particles the LHC was built to observe [34, 35]. Therefore, searching for signs of these more complex theories is generally more challenging and requires dedicated techniques to identify and reconstruct candidate particles that would hint at the presence of dark matter. These searches are now at the forefront of the ATLAS and LHC quest for dark matter, and have gathered at least as much interest as searches for WIMPs and their associated particles.

Searching for non-WIMP dark matter is well-motivated but generally more challenging than MET+X searches, as it requires dedicated techniques to identify and reconstruct candidate particles that would hint at the presence of dark matter.

One example search out of many, highlighted because of its connection with searches in the LHCb experiment (discussed in the relevant chapter of this book), is looking for a portal particle akin to those mentioned in Sect. 4.2, called dark boson or dark photon. This dark photon can be likened to a massive equivalent of the Standard Model photon (with which it interacts). It arises from a new dark force that mirrors the electromagnetic force we know, but this new dark force is much weaker. Many target models sought by ATLAS also contain other particles in addition to the dark photon, so care is needed when comparing results with other searches looking for minimal models containing dark photons only (as in the case of some of the LHCb searches mentioned in this book).

In some theoretical models, dark photons are radiated from new dark fermions and produced in association with other new particles that under certain circumstances

could serve as dark matter candidates, or can arise from the decays of dark Higgs bosons [36]. ATLAS looks for the decay products of the dark photons predicted by these models, for example Standard Model leptons [37]. Since the dark photons have low masses (order of 1 GeV) and are produced from the decays of massive particles, their decay products are Lorentz-boosted and are collimated in the same detector area (in the case of decays into leptons, these are called lepton-jets). This is a useful feature for enhancing signal over background, since the other Standard Model processes that produce pairs of leptons very close to each other are relatively rare and can be suppressed further.

Searches involving even weaker interactions between the dark photon and the regular photon, and in general with very weakly coupled particles, lead to interesting features in terms of detector signatures. For example, the dark photon decay products can be displaced with respect to the collision vertex due to the long lifetime of the dark photon. This feature can be exploited in a targeted search for displaced lepton-jets [38].

It's also important to note that general-purpose experiments are only one of the many ways to search for dark photons and for other long-lived and feebly interacting particles belonging to dark sectors at colliders - for more, see e.g. Ref. [39] and the BASE, FASER, CAST and NA64 chapters. Often there is a trade-off between being able to search for the largest possible number of interactions and detector signatures and being able to target specific models with unusual detector signatures with the best possible sensitivity. This means that both approaches, both general-purpose and specialised detectors, are needed for a broad coverage of dark matter hypotheses at colliders.

5 Connecting ATLAS Searches to Astrophysical Searches in WIMP Scenarios

Searches for dark matter at the LHC are typically searches for the production, rather than the interaction or annihilation, of potential dark matter particles. As such, data from ATLAS would not provide proof that a new particle constitutes the dark matter - the sensitivity to dark matter lifetimes is just too short (see above). Nevertheless, ATLAS data could establish consistency with the predictions of dark matter models, and within those models ATLAS can provide complementary information to a broad range of astroparticle searches for the interaction of relic WIMP dark matter particles being carried out around the world.

This complementarity can be illustrated taking as example the simple dark matter-mediator model illustrated in Fig. 6.

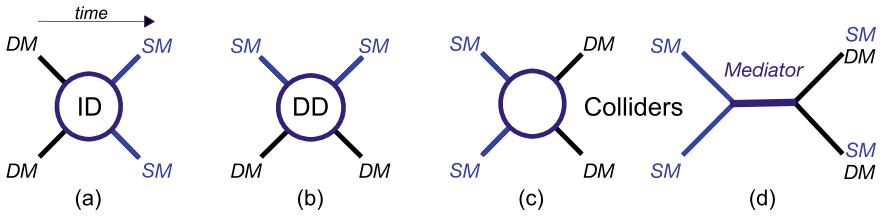

Fig. 11 Schematic illustration of Dark Matter (DM) interactions and their corresponding experimental detection techniques, with time flowing from left to right. Figure **a** shows DM annihilation to Standard Model (SM) particles, as sought by Indirect Detection (ID) experiments. Figure **b** shows DM -> SM particle scattering, targeted by Direct Detection (DD) experiments. Figure **c** shows the production of DM particles from the annihilation of SM particles at colliders. Figure **d** shows again the pair production of DM at colliders as in **c**, but in this case the interaction occurs through a mediator particle between DM and SM particles. *Source* and ©: Wikimedia commons, C. Doglioni and A. Boveia

ATLAS can provide complementary information to a broad range of astroparticle searches for the interaction of relic WIMP dark matter particles being carried out around the world.

Within this model, in order for dark matter particles to be produced in pairs at the LHC, two strongly interacting quarks or gluons from the colliding protons must interact to produce the two dark matter particles (Fig. 11b). These same interactions could enable relic dark matter particles trapped in the Milky Way galaxy to scatter off atomic nuclei on Earth, generating the nuclear recoil signature exploited by "direct" astroparticle searches for dark matter such as XENON in Europe, LUX in North America, PANDA-X in China and IceCube at the South Pole (see the chapter about underground searches in this book). Constraints from ATLAS searches can therefore be translated, albeit with assumptions on the mediator-proton and mediator-dark matter interaction, into constraints on the possible signals in those experiments. Figure 12 shows that recent ATLAS searches and astroparticle searches for dark matter have a complementary reach when plotted in terms of interactions with nucleons (y axis) and dark matter mass (x axis), in the case of the simple dark matter-mediator model of Fig. 6.

Furthermore, the same interactions also enable relic dark matter particles produced in the early Universe to annihilate and create Standard Model particles (Fig. 11a). This leads to the signatures for dark matter sought by "indirect" dark matter search experiments - typically high-energy photons (observed by telescopes such as HESS, MAGIC and VERITAS), neutrinos (observed by neutrino telescopes such as IceCube) or anti-particles (detected by space experiments such as AMS on the International Space Station, described in this book). Results from collider searches can therefore also be compared with results from those experiments.

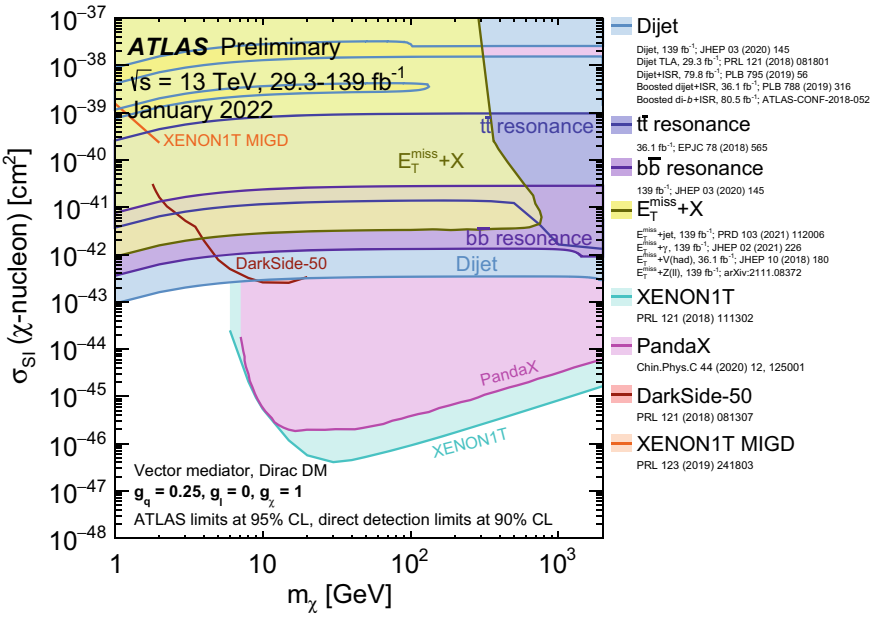

Fig. 12 A comparison of the inferred limits from ATLAS data, including those from both MET+X and mediator resonance searches, to the constraints from direct detection experiments on the WIMP-proton scattering cross section in the context of a model with a new vector particle mediating the standard model - dark matter interaction, fixing the given mediator-quark (g_q) and mediator-dark matter (g_{DM}) coupling strengths to the value in the plot. Figure from [40], ©CERN

It is important to note that the plots as in Fig. 12 only provide a single, limited snapshot of the complementarity between collider experiments and direct detection experiments. This is because in the results from collider searches the interaction strengths between mediator, dark matter and Standard Model particles (noted as g_DM, g_q and g_l in Fig. 12), as well as the kind of dark matter and mediator particles, are fixed. We don't know whether those exact couplings and particles are realised in nature. For example, if the interaction strength between quarks and mediator was much weaker than the value chosen for displaying results, the LHC constraints on this kind of model would also become weaker. As pointed out in the last section of this chapter, work is ongoing to display cases of weaker-coupled mediators as well, also to be able to compare collider results with other experiment at accelerators - for example, see the NA64 and FASER chapters of this book.

Moreover, these plots do not take into account other astrophysical or theoretical uncertainties affecting any kind of experiments.

5.1 Considerations on Dark Matter Relic Density for the Comparison of Astrophysics Searches and Collider Searches

When interpreting and combining ATLAS results and those from astroparticle dark matter searches, we can also consider whether the dark matter model being tested is consistent with the observed density of relic dark matter particles (note that this is not done for the ATLAS results in Fig. 13, while it is an assumption that enters the WIMP-proton scattering cross section). This relic density has been measured with a precision better than 1% through observations of the cosmic microwave background by satellites such as Planck [41]. When considering a particular dark matter model, this can help setting an upper limit on the amount of dark matter the model should produce. This is because, in principle, the dark matter could consist of multiple types of particles, with any one type only contributing a fraction of the measured amount. However, other interaction mechanisms that are not considered in the particular model but are still compatible with it could also deplete the dark matter abundance, motivating searches that are only loosely tied to the relic density target. The relic dark matter density constraint is considered particularly important for more complete (e.g. SUSY) dark matter models, given that they are consistent models that can

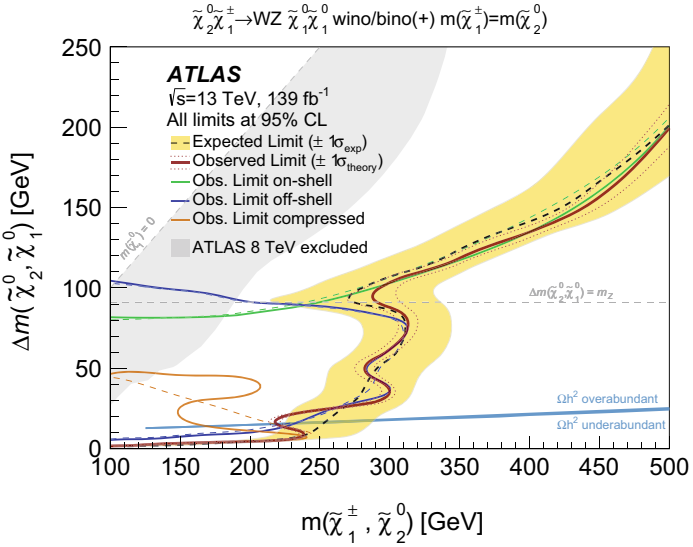

Fig. 13 Exclusion limits (shown as continuous and dashed lines) in terms of SUSY particle masses and mass differences between the two particles for a specific SUSY model. Certain light particles predicted by SUSY are able to give us the amount of dark matter abundance equal to the observed relic density, given the set of SUSY parameter models described in the paper. The area above (below) the blue line represents a dark-matter relic density larger (smaller) than the observed one. Figure from [42], ©CERN

provide answers for multiple questions in the Standard Model of particle physics. These models can often predict more dark matter than the Planck satellite observed. Special characteristics of the model, such as closely-spaced SUSY particle masses or increased dark matter interactions, can reduce this density to values consistent with observations, and searches for models with these characteristics are a high priority for ATLAS. An example of a plot displaying the relic density and pointing the way to parameter space where the model sought would produce the observed abundance is shown in Fig. 13.

6 Outlook: Where Do We Go From Here?

- ATLAS is searching for dark matter at the LHC in synergy with other experimental collaborations, such as CMS and LHCb. LHC experiments haven't yet discovered dark matter candidates but there is a large number of proton-proton collisions ahead to make the most of. The upcoming LHC data-taking period (2022–2024, known as Run 3) is expected to more than double the current dataset, and the high-luminosity period beginning 2027 will deliver at least another factor of 10 more data. The experiments will be able to probe dark matter processes that are rarer and more challenging to reconstruct than the ones studied today. In view of the upcoming data-taking, experiments are also making use of more advanced data collection and data analysis techniques, such as machine learning.[4]
- Direct and indirect searches for signals of the existing dark matter in our galactic neighbourhood are important complementary strategies to LHC searches, since astrophysical experiments are able to detect relic dark matter and they are necessary to confirm that a new invisible particle discovered at the LHC could make up dark matter. We will continue the dialogue with these experiments, exchanging scientific results and perspectives, share theoretical models, and extend the discussion to the broader astrophysics community. A number of cross-experiment and cross-field initiatives have been created for this purpose. For example, the Initiative for Dark Matter in Europe and Beyond (iDMEu) is a joint initiative supported by the European communities for accelerator, astroparticle and nuclear physics (JENAA) that works towards this purpose (see http://www.idmeu.org).
- Many other beyond-collider experiments can probe dark matter models to which the LHC experiments are not sensitive, for example models where the interactions between dark matter and ordinary matter are too feeble for dark matter to be produced in collisions of known particles. These experiments are being discussed in the Physics Beyond Colliders effort at CERN (see https://pbc.web.cern.ch and [39]).

[4] For more information on ongoing collaborative efforts on Machine Learning in Dark Matter, see the DarkMachines research collective webpage: http://darkmachines.org/. For general perspectives on data acquisition and collection see the HEP Software Foundation webpage: https://hepsoftwarefoundation.org/.

- As one of the main outstanding questions in fundamental physics, the identification of the nature of dark matter is a key scientific driver for the future of particle physics. For this reason dark matter searches are a main focus of the discussions, including both experimentalists and theorists, which have taken place in recent initiatives to draw up roadmaps for the future of the field (see [43] and https://snowmass21.org). While the nature of dark matter is currently still unknown, it is clear that the quest to better understand it will be a highlight of humanity's study of the fundamental constituents of the universe for many years to come.

Acknowledgements Research by C. D. is part of projects that have received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement 679305 and 101002463) and from the Swedish Research Council. Research by D. T. is part of a project that has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement 694202) and from the UK Science and Technology Facilities Council (STFC).
We thank the following collaborators for useful input and review (including on the original feature article): Katarina Anthony, Antonio Boveia, Oleg Brandt, Carl Gwilliam, Andreas Hoecker, Marie-Helene Genest, Suchita Kulkarni, Zachary Marshall, Christian Ohm, Giordon Stark.

References

1. D. Caterina, T. Dan, Searching for Dark Matter with the ATLAS detector (2019). https://atlas.cern/updates/feature/dark-matter
2. N. Bock, Antiproton discovery (2009). https://www.symmetrymagazine.org/article/october-2009/antiproton-discovery
3. ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. **B716**, 1–29 (2012). https://doi.org/10.1016/j.physletb.2012.08.020
4. G. Heather, M. Bruno, The Higgs boson: the hunt, the discovery, the study and some future perspectives (2018). https://atlas.cern/updates/feature/higgs-boson
5. L. Evans, P. Bryant, LHC machine. JINST **3**, S08001 (2008)
6. CERN, The Large Hadron Collider (2020). https://home.cern/science/accelerators/large-hadron-collider
7. ATLAS Collaboration, The ATLAS experiment at the CERN large hadron collider. JINST **3**, S08003 (2008). https://doi.org/10.1088/1748-0221/3/08/S08003
8. ATLAS Collaboration, Discover the ATLAS detector (2018). https://atlas.cern/discover/detector. https://atlas.cern/updates/feature/higgs-boson
9. J. Pequenao, Computer generated image of the whole ATLAS detector (2008). https://cds.cern.ch/record/1095924
10. LHC Dark Matter Working Group, LHC Physics Centre Working Group on Dark Matter Searches at the LHC (2015). http://lpcc.web.cern.ch/content/lhc-dm-wg-wg-dark-matter-searches-lhc
11. D. Abercrombie, N. Achkurin, E. Akilli, J. Alcaraz Maestre et al., Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum (2015)
12. A. DiFranzo, K.I. Nagao, A. Rajaraman, T.M.P. Tait, Simplified models for dark matter interacting with quarks. JHEP **11**, 014 (2013). https://doi.org/10.1007/JHEP11(2013)014https://doi.org/10.1007/JHEP01(2014)162. [Erratum: JHEP01,162 (2014)]
13. O. Buchmueller, M.J. Dolan, C. McCabe, Beyond effective field theory for dark matter searches at the LHC. JHEP **01**, 025 (2014). https://doi.org/10.1007/JHEP01(2014)025

14. M. Chala, F. Kahlhoefer, M. McCullough, G. Nardini, K. Schmidt-Hoberg, Constraining dark sectors with monojets and dijets (2015)
15. J. Miguel No, Looking through the pseudoscalar portal into dark matter: novel mono-Higgs and mono-Z signatures at the LHC. Phys. Rev. D **93**(3), 031701 (2016). https://doi.org/10.1103/PhysRevD.93.031701
16. T. Abe et al., LHC dark matter working group: next-generation spin-0 dark matter models. Phys. Dark Univ. **27**, 100351 (2020). https://doi.org/10.1016/j.dark.2019.100351
17. G. Aad et al., Search for associated production of a Z boson with an invisibly decaying Higgs boson or dark matter candidates at $\sqrt{s} = 13$ TeV with the ATLAS detector (2021)
18. M. Aaboud et al., Measurement of the $Z\gamma \to \nu\bar{\nu}\gamma$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector and limits on anomalous triple gauge-boson couplings. JHEP **12**, 010 (2018). https://doi.org/10.1007/JHEP12(2018)010
19. G. Aad et al., Search for dark matter in association with an energetic photon in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. JHEP **02**, 226 (2021). https://doi.org/10.1007/JHEP02(2021)226
20. G. Aad et al., Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector. Phys. Rev. D **103**(11), 112006 (2021). https://doi.org/10.1103/PhysRevD.103.112006
21. ATLAS Collaboration, Jetting into the dark side: a precision search for dark matter (2020). https://atlas.cern/updates/briefing/precision-search-dark-matter
22. G. Aad et al., Observation of electroweak production of two jets in association with an isolated photon and missing transverse momentum, and search for a Higgs boson decaying into invisible particles at 13 TeV with the ATLAS detector (2021)
23. M. Aaboud et al., Search for dark matter in events with a hadronically decaying vector boson and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. JHEP **10**, 180 (2018). https://doi.org/10.1007/JHEP10(2018)180
24. G. Aad et al., Search for new phenomena in events with two opposite-charge leptons, jets and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. JHEP **04**, 165 (2021). https://doi.org/10.1007/JHEP04(2021)165
25. ATLAS Collaboration, The supersymmetric bottom quark and its friends (2021). https://atlas.cern/updates/briefing/supersymmetric-bottom-quark-friends
26. G. Aad et al., Search for dark matter produced in association with a single top quark in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. Eur. Phys. J. C **81**, 860 (2021). https://doi.org/10.1140/epjc/s10052-021-09566-y
27. M. Aaboud et al., Search for low-mass dijet resonances using trigger-level jets with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV. Phys. Rev. Lett. **121**(8), 081801 (2018). https://doi.org/10.1103/PhysRevLett.121.081801
28. ATLAS Collaboration, A new data-collection method for ATLAS aids in the hunt for new physics (2018). https://atlas.cern/updates/briefing/new-data-collection-method-atlas-aids-hunt-new-physics
29. M. Aaboud et al., Constraints on mediator-based dark matter and scalar dark energy models using $\sqrt{s} = 13$ TeV pp collision data collected by the ATLAS detector. JHEP **05**, 142 (2019). https://doi.org/10.1007/JHEP05(2019)142
30. ATLAS Collaboration, ATLAS releases new results in search for weakly-interacting supersymmetric particles (2017). https://atlas.cern/updates/briefing/atlas-releases-new-results-search-weakly-interacting-supersymmetric
31. M. Aaboud et al., Search for electroweak production of supersymmetric particles in final states with two or three leptons at $\sqrt{s} = 13$ TeV with the ATLAS detector. Eur. Phys. J. C **78**(12), 995 (2018). https://doi.org/10.1140/epjc/s10052-018-6423-7
32. R. George, de Jong Paul, Broken symmetry: searches for supersymmetry at the LHC (2018). https://atlas.cern/updates/feature/supersymmetry
33. ATLAS Collaboration, The hunt for higgsinos reaches new limits (2021). https://atlas.cern/updates/briefing/new-higgsino-limits

34. J. Alimena et al., Searching for long-lived particles beyond the Standard Model at the Large Hadron Collider. J. Phys. G **47**(9), 090501 (2020). https://doi.org/10.1088/1361-6471/ab4574
35. L. Lee, C. Ohm, A. Soffer, Yu. Tien-Tien, Collider searches for long-lived particles beyond the standard model. Prog. Part. Nucl. Phys. **106**, 210–255 (2019). https://doi.org/10.1016/j.ppnp.2019.02.006
36. M. Buschmann, J. Kopp, J. Liu, P.A.N. Machado, Lepton jets from radiating dark matter. JHEP **07**, 045 (2015). https://doi.org/10.1007/JHEP07(2015)045
37. G. Aad et al., A search for prompt lepton-jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. JHEP **02**, 062 (2016). https://doi.org/10.1007/JHEP02(2016)062
38. G. Aad et al., Search for long-lived neutral particles decaying into lepton jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. JHEP **11**, 088 (2014). https://doi.org/10.1007/JHEP11(2014)088
39. J. Beacham et al., Physics beyond colliders at CERN: beyond the standard model working group report. J. Phys. G **47**(1), 010501 (2020). https://doi.org/10.1088/1361-6471/ab4cd2
40. ATLAS Collaboration, Summary plots from the ATLAS Exotic physics group (2017). https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/EXOTICS/index.html
41. P.A.R. Ade et al., Planck 2015 results. XIII. Cosmological parameters. Astron. Astrophys. **594**, A13 (2016). https://doi.org/10.1051/0004-6361/201525830
42. G. Aad et al., Search for chargino-neutralino pair production in final states with three leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. Eur. Phys. J. C **81**, 1118 (2021). https://doi.org/10.1140/epjc/s10052-021-09749-7
43. R. Keith Ellis et al., Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020 (2019)

Searching for Dark Matter with the CMS Detector

Deborah Pinna

1 Introduction

As explained in previous chapters, proof of the existence of dark matter (DM) in our universe is provided by various astrophysical observations [1–3], but its nature or its non-gravitational interactions still remain elusive. One of the most studied class of theories assumes dark matter to be a stable, weakly interacting massive particles, and these translates in the possibility to produce dark matter in high energy collisions at the LHC [4].

If produced in the collision, the dark matter particles will not leave any direct traces in the detector (such as CMS [5]), but their presence can be inferred by a large transverse momentum imbalance (so called MET or p_T^{miss}) if produced in association with a detectable physics object. These type of searches are usually referred as mono-X searches, where X identifies the visible object.

Most of the searches for dark matter performed at CMS have a common strategy, which exploit the presence of a highly energetic object X in association with a large missing transverse energy in the collision event. This signature can also appear in events from Standard Model (SM) processes (referred to as "backgrounds"), therefore the dark matter production will appear as an excess of events in the tail of the MET distribution over the Standard Model expectations. Unfortunately, this is not a striking signature such as a mass peak or a kinematic end point, making dark matter searches very challenging. It is therefore essential to have a precise modeling and evaluation of the background processes in the phase space enriched in signal events (signal region - SR), which is achieved through dedicated control regions (CR) enriched in the background of interest. The results are then extracted comparing the Standard Model predictions to data and if no significant excess is found the results are interpreted in terms of the model parameters that predict the dark matter production at colliders.

D. Pinna (✉)
University of Wisconsin-Madisonn, Madison, WI, USA
e-mail: deborah.pinna@cern.ch

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_6

This type of search strategy allows to make very general assumptions on the nature of the dark matter particles, i.e. they can be produced in the proton-proton collision since they interact with Standard Model particles and they will leave no direct trace in the CMS detector. Therefore, the results of the analyses can be interpreted under a variety of dark matter models. A convenient theoretical framework to interpret the results is provided by the so-called simplified models [6], where the dark matter particles are assumed to be produced from the decay of a new mediator that interacts (or couples) both to dark matter and Standard Model particles. Different assumptions on this mediator can be investigated at colliders, for e.g. a spin-0 or a spin-1 particle leading to scalar/pseudoscalar or vector/axial vector interaction respectively. As explained in previous chapters, these models are an appealing benchmark for dark matter searches at LHC because of the limited number of parameters (dark matter mass m_{DM}, mediator mass and couplings g_q and g_χ) that enclose the relevant physics process features [6].

As already mentioned in previous chapters, many challenges characterize searches for dark matter at colliders, as an e.g. an accurate energy calibration and resolution of the visible objects is essential as well as a precise particle reconstruction and identification. For these reasons, the various analyses performed at CMS employ dedicated techniques to efficiently identify and reconstruct visible particles and dark matter candidate, as well as using advanced data-collection and data-analysis techniques, such as machine learning, or non-standard data recording techniques. In the following, a selection of various type of searches performed at CMS using the data collected from 2016 until 2018 are presented to highlight some of the different techniques and theoretical frameworks employed in the hunt for dark matter.

The results of the various searches presented provide complementary information on different parameters of dark matter models and these different approaches could play a major role in understanding the nature of dark matter in case of discovery. Many models employed make very simple assumptions on the dark matter interactions with Standard Model particles, while these are very general the true nature of these particles might be more complex. For this reason, in addition to the presented searches, the CMS collaboration targets many other experimental signatures for the hunt of dark matter, as extended dark sectors, to be able to cover a large variety of regions in the model parameter space that are not yet covered. A comprehensive summary of searches for dark matter at CMS can be found in the public results page [7].

2 Simplified Models and Extensions

Simplified models provide a tantalizing possibility for dark matter discovery at CMS, where dark matter interactions with Standard Model particles are assumed to be mediated by a new particle. Different assumptions on this mediator can be made, for example it can be assumed to be a spin-1 particle leading to vector/axial vector interactions. These interactions, for Dirac fermion dark matter particles produced

via the s-channel exchange of a spin-1 mediator Z', are described by the following Lagrangians [6]:

$$\mathcal{L}_{\text{vector}} \supset g_q \sum_q Z'_\mu \bar{q}\gamma^\mu q, \quad \mathcal{L}_{\text{axial-vector}} \supset g_q \sum_q Z'_\mu \bar{q}\gamma^\mu\gamma^5 q \tag{1}$$

where the coupling g_q is assumed to be equal for all quarks. From these formulas, one can infer that to increase the discovery potential of dark matter from these interactions events where the choice of the detectable physics object leads to an increase of the production cross section or the background rejection should be favored. This has therefore motivated analyses searching for events in which the dark matter particles are produced in association with a jet or vector bosons V (Z or W boson) [8].

Another theoretically attractive possibility are interactions mediated by a new scalar (ϕ) or pseudoscalar (A) particle, as they can be easily accommodated in extended Higgs sectors [9, 10]. Assuming minimal flavor violation [11–14], the couplings between the neutral spin-0 mediator and the Standard Model particles are proportional to the quark masses m_q [6]:

$$\mathcal{L}_{\text{scalar}} \supset g_q \frac{\phi}{\sqrt{2}} \sum_q y_q \bar{q}q, \quad \mathcal{L}_{\text{pseudoscalar}} \supset g_q \frac{iA}{\sqrt{2}} \sum_q (y_q \gamma^5 q) \tag{2}$$

where $y_q = \sqrt{2}m_q/v$ are the Yukawa couplings with v the Higgs vev, g_q is the fermion-mediator coupling assuming a universal value for all quark q flavors. In this case, the discovery potential can be increased considering events where the visible object increases the production cross section as in mono-jet processes, or exploits the proportionality of the coupling of the mediator to heavier quarks as done in analyses searching for events in which the dark matter particles are produced in association with top quarks [15].

One analysis performed by the CMS collaboration that investigated these interactions considers hadronically-decaying high-p_T V bosons or jets recoiling against large MET [8]. A representative Feynman diagram for mono-jet events with a spin-1 Z' mediator which decays into two dark matter particles χ is presented in Fig. 1 [left]. Events are required to have large MET and no leptons and are categorized based on the nature of the jet. If the vector boson is very energetic (large Lorentz boost in the laboratory frame) it will be reconstructed as a single large-radius jet and selection requirements can be applied to guarantee the consistency with hadronic V decays based on the large-radius jet invariant mass and substructure quantities [16]. If the jet is not consistent with hadronic V decays is classified as a mono-jet event.

For energetic V bosons, the hadronic decay products are Lorentz boosted in the laboratory frame and are reconstructed as a single large-radius jet with a characteristic substructure. Machine learning algorithms based on artificial neural networks are used in order to identify such signatures and efficiently suppress the over-whelming background coming from quantum chromodynamics (QCD) production of jets [14].

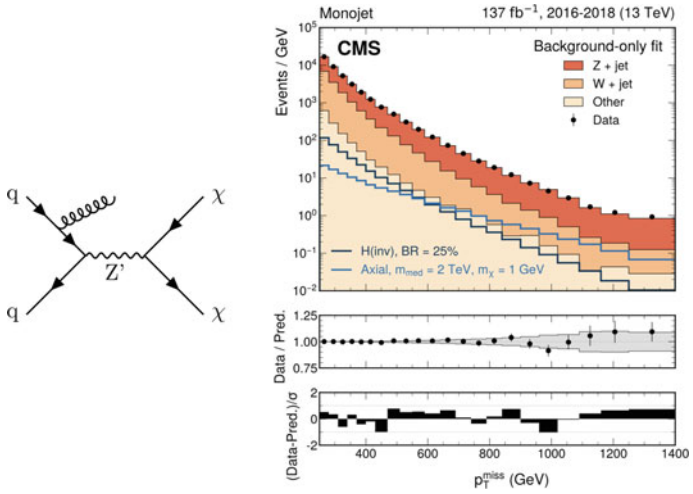

Fig. 1 [left] Representative Feynman diagram for mono-jet events with a spin-1 Z' mediator which decays into two dark matter particles χ [8], [right] MET distribution for mono-jet events. The solid histograms for the simulated Standard Model backgrounds are summed cumulatively. The data are represented by solid points on which the horizontal bar indicates the width of the bin. In the middle plots the ratio between data and the total Standard Model background is presented with the associated uncertainty shown by the grey band [8]

The main backgrounds consist of $t\bar{t}$, $Z(\nu\nu)$, and $W(l\nu)$ + jets events, which are estimated from data in dedicated CRs. The results are then extracted by a simultaneous fit in SRs and CRs to the MET distribution, where the systematic uncertainties are included as nuisance parameters. No significant excess above SM expectation is observed as can be seen in Fig. 1 [right] where the MET distribution is shown for events classified as mono-jet. The results are interpreted in terms of benchmark simplified models where the only new particles are the s-channel mediator and the dark matter particles (simplified pseudoscalar or axial-vector model), or a colored t-channel mediator decaying into one dark matter particle and a quark (fermion portal). The corresponding 95% confidence level (CL) limits on the production cross section as a function of the mediator mass and of the dark matter mass for benchmark values of the other model couplings are presented in Fig. 2. In the case of a pseudoscalar mediator, the blue solid line indicates the exclusion boundary and mediator masses values up to 470 GeV are excluded. For the axial-vector mediator model, parameter combinations within the solid black line are excluded, reaching up to mediator masses of 1.95 TeV for low values of dark matter mass ($m_{DM} = 1$ GeV). The gray dashed line indicates the diagonal where the mass of the mediator m_{med} had values such that $m_{med} = 2m_{DM}$, above which only off-shell mediator production contributes to the mono-jet and mono-V final states. Similarly, for the fermion portal model at low m_{DM} values, mediator masses up to 1.5 TeV are excluded, which are the most stringent constraints on this model to date.

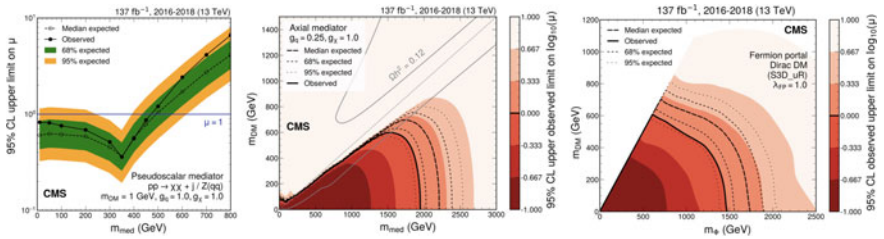

Fig. 2 [left] Upper limits at 95% CL on the ratio μ (y-axis) between the excluded cross section and its theoretical value as a function of the mediator mass m_{med} (x-axis) for benchmark values of the couplings $g_q = 1.0$ and $g_q = 1.0 and g_\chi = 1.0$ for a spin-0 pseudoscalar mediator. The blue solid line indicates the exclusion boundary $\mu = 1$ [8], [center] Exclusion limits at 95% CL on the ratio μ between the excluded cross section and its theoretical value as a function of the mediator mass m_{med} (x-axis) and of the dark matter mass m_{DM} (y-axis) for benchmark values of the couplings $g_q = 0.25$ and $g_\chi = 1.0$ for a spin-1 axial-vector mediator. The black solid line indicates the observed exclusion boundary for $\mu = 1$ and all the parameters values enclosed inside this line are excluded by this search [8], [right] Exclusion limits at 95% CL in the plane of the mediator mass m_Φ and the dark matter mass m_{DM} in the fermion portal model. The black solid line indicates the observed exclusion boundary and all the parameters values enclosed inside this line are excluded by this search [8]

The constraints placed on the simplified models imply constraints also on the interaction cross section between dark matter and nuclei. The results can in fact be translated into upper bounds on the dark matter-nucleus scattering cross section as a function of the dark matter mass m_{DM}. Depending on the mediator type, the resulting couplings are either spin dependent (axial-vector) or independent (vector) as shown in Fig. 3. It can be inferred that, under the assumptions made, collider results depend weakly on m_{DM} in the on-shell regime, leading to more stringent constraints at low values of m_{DM} compared to astroparticle searches. This is also an example of the importance of the complementarity between collider and astroparticle searches for the discovery of dark matter.

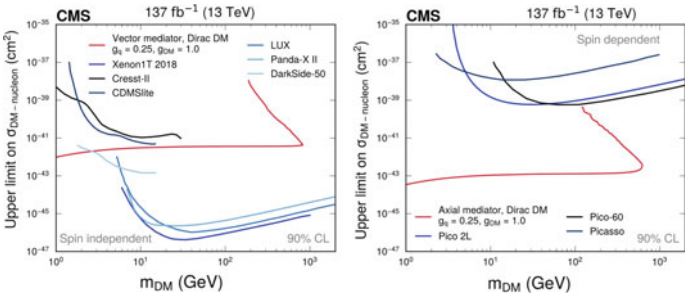

Fig. 3 Comparison of the simplified model constraints (red line) to results from direct-detection experiments (blue lines). The comparison is shown separately for the vector [left] and axial-vector [right] mediators, which translate into spin-independent and spin-dependent dark matter-nucleon couplings, respectively [8]

3 The Invisible Through the Visible

As seen in the previous section, events where the dark matter particles are produced in association with a jet or vector bosons (Z or W) can be investigated to study vector/axial-vector interactions mediated by a new spin-1 particle which decays into dark matter particles. From the results presented in Fig. 2 [center], it is evident that these type of searches are sensitive in the region where the assumed mediator mass is higher than two times the dark matter mass (so called on-shell regime). Instead, in the off-shell region the sensitivity decreases drastically due the strong reduction of the production cross section.

Since the new mediator is assumed to interact both to dark matter and Standard Model particles, it could also decay to visible particles. This type of events will appear as a peak (or "resonance") above the Standard Model background in the distribution of the total mass of the visible particles, and can provide complementarity to mono-X searches in the off-shell regime.

The energies reached by the LHC allow to search for high-mass resonances ($\mathcal{O}$(TeV)). Nevertheless, dark matter mediators could have masses which are lower than this, but such region is very difficult to probe due to the overwhelming background. The solution employed in analyses performed at CMS to probe lower mediator mass are various techniques such as selecting visible objects that allow lower trigger thresholds and/or advanced data-taking techniques that allow more events to be recorded at the cost of retaining smaller amount of information for each event.

One of these analyses performed at CMS considers events with a pair of electrons or muons where the dark matter mediator appears as a peak with respect to the background in the invariant mass spectrum of the two leptons [17]. The main backgrounds consist of $t\bar{t}$, Z(ll), and multi-jets and W + jets events where the leptons are mis-identified, and are all estimated from data in dedicated CRs. This search can investigate lower masses with respect to the classic searches for di-jet resonances thanks to the lower thresholds for lepton triggers. The results are then extracted by a fit to the dilepton invariant mass. An example of the invariant mass distribution for events where an electron pair is selected is shown in Fig. 4, and no significant excess above background expectations is observed.

The results are then interpreted in terms of benchmark simplified models where the only new particles are the s-channel mediator and the dark matter particles (simplified vector and axial-vector models). The corresponding 95% confidence level upper limits are presented in Fig. 5. The limits are strongest for large values of dark matter and mediators masses below 1.92 (4.64) TeV are excluded in the vector (axial-vector) model.

In order to probe even lower masses ($\mathcal{O}$(GeV)) for the dark matter mediator, only trigger-level objects information can be retained during data-taking. This technique is employed by a CMS analysis looking for dimuon low-mass resonances [18]. Events with an opposite-sign muon pair are selected employing a dedicated dimuon trigger with low p_T threshold and high rate. The 4-momentum, isolation and track of the muons is the only information retained for each event. As can be seen from Fig. 6,

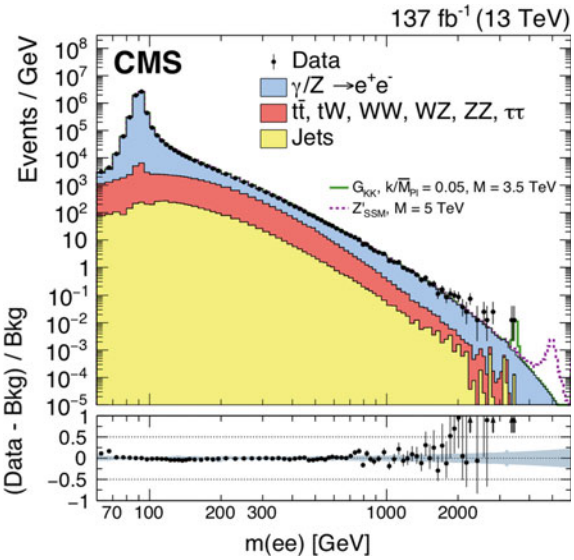

Fig. 4 Invariant mass distribution of pairs of electrons. The solid histograms for the simulated Standard Model backgrounds are summed cumulatively. The data are represented by solid black points with statistical uncertainties. The ratios of the data yields after background subtraction to the expected background yields are shown in the lower plots. The blue shaded band represents the combined statistical and systematic uncertainties in the background [17]

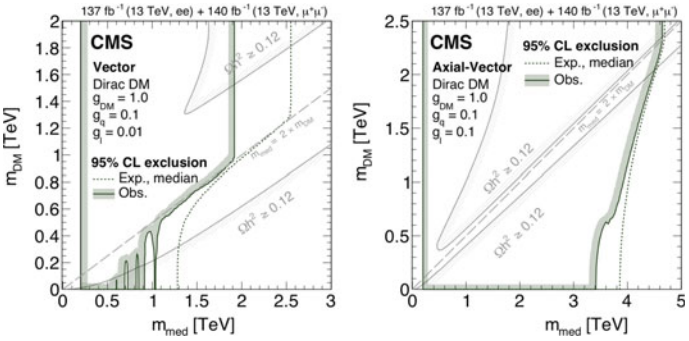

Fig. 5 Upper limits at 95% CL on the production cross section as a function of the dark matter mass (y-axis) and on its associated mediator, in a simplified model of DM production via a [left] vector or [right] axial-vector mediator (x-axis) for benchmark values of the couplings $g_q = 0.1$ and $g_\chi = 1.0$. The curves with the hatching represent the excluded parameters regions [17]

where the distribution of the dimuon mass is shown, masses as low as few GeV can be probed with these data-taking techniques. The main backgrounds consist of events where the dimuon vertex is wrongly associated, or events with prompt muons or from material vertices.

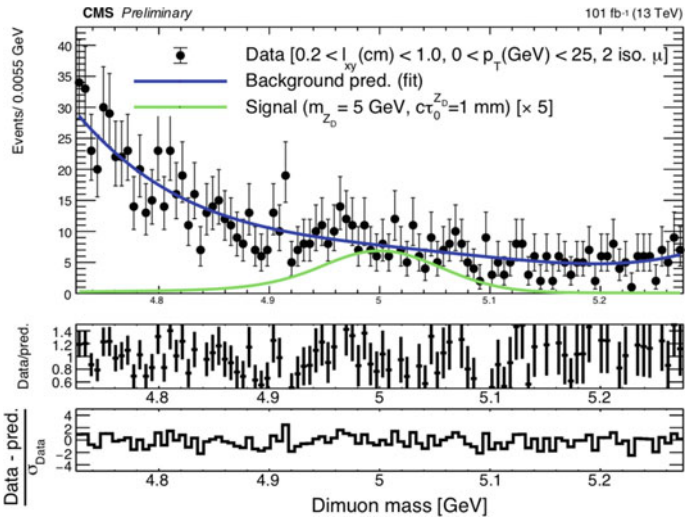

Fig. 6 Dimuon invariant mass distribution. The solid histograms for the simulated Standard Model backgrounds are summed cumulatively. The data are represented by solid black points with statistical uncertainties. The ratios of the data yields after background subtraction to the expected background yields are shown in the lower plots [18]

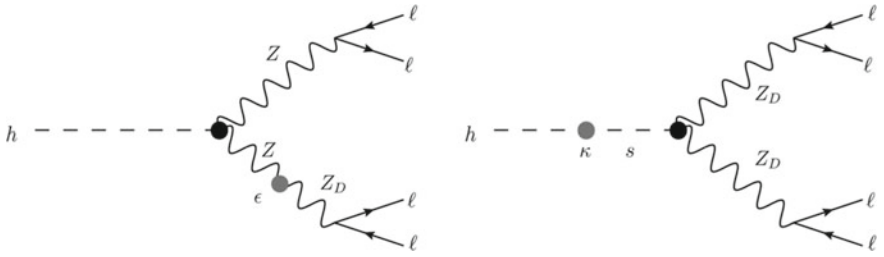

Fig. 7 Representative Feynman diagrams illustrating a Standard Model-like Higgs boson (h) decay to four leptons via one or two intermediate Z_D [19] through the hypercharge portal [left] or through the Higgs portal [right] [18]

No significant excess above SM expectation is observed. The results then interpreted in terms of benchmark models where dark and Standard Model sectors interact through a new particle, a dark photon Z_D which can decay into Standard Model particles [19]. Representative Feynman diagrams for these type of events are presented in Fig. 7. The corresponding 95% confidence level upper limits on the branching fraction $B(h \rightarrow Z_D Z_D)$ as function of the dark photon mass m_{Z_D} are presented in Fig. 8.

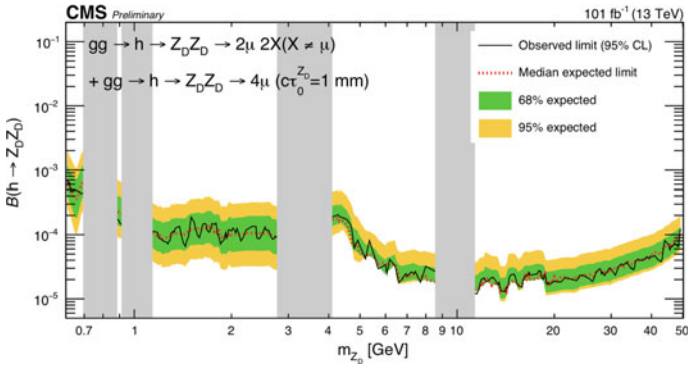

Fig. 8 Upper limits at 95% CL on the branching fraction $B(h \rightarrow Z_D Z_D)$ as a function of the dark photon mass m_{Z_D} (x-axis). The black solid line indicates the observed exclusion boundary, while the vertical gray bands indicate mass ranges containing known Standard Model resonances, which are masked in the search [18]

From the presented searches, it can be seen that collider signatures beyond classical mono-X events can provide important complementarity in the hunt for dark matter and essential constraints on various dark matter models.

4 Higgs Boson: Extended Sectors and Invisible Decays

4.1 Two-Higgs Doublet Extensions

The Higgs boson discovery [20, 21] provides a new collider probe for dark matter at colliders. In fact, it can be considered as the detectable physics object X leading to mono-H signatures. The CMS collaboration investigates mono-H events for dark matter discovery considering different H decay modes [22]. The most sensitive channel is provided by the H decay to two b quarks due to the high branching ratio, while decays to V boson pairs provide more clean signatures and therefore higher background rejection despite the lower branching ratio.

The Higgs boson also offers a probe for dark matter in extended dark sectors. In this scenario, the dark matter particle acquires mass through its interaction with a new particle, the dark Higgs boson s [23]. The mono-H signature in this case is produced by a spin-1 mediator Z' (as in the simplified models introduced in previous sections) or by a dark matter intermediate state χ as shown in the Feynman diagrams in Fig. 9. In this context, the additional parameters of the model include the dark Higgs mass m_s and the mixing angle between Standard Model and the dark Higgs bosons ($\sin\theta$). A search for this type of signature is performed by CMS for events where the dark Higgs boson decays to WW bosons in final states with two leptons [24]. The WW channel is the most relevant decay mode in terms of cross section

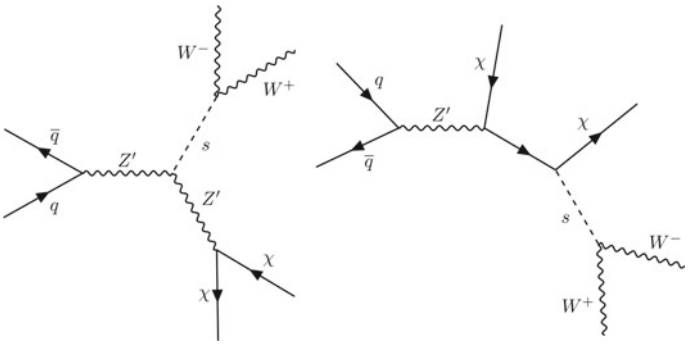

Fig. 9 Representative Feynman diagrams illustrating mono-H events produced by a spin-1 mediator Z' or by a dark matter intermediate state χ, where the dark Higgs boson decays to a WW vector boson pair [24]

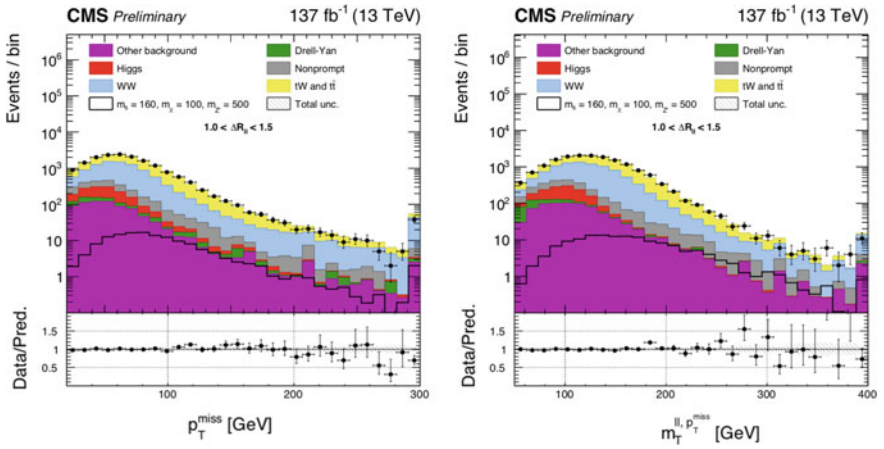

Fig. 10 Distributions for the missing transverse momentum p_T^{miss} [left] and for the transverse mass of the dilepton system plus p_T^{miss}, $m_T{}^{ll,p_T^{miss}}$ [right]. The solid histograms for the simulated Standard Model backgrounds are summed cumulatively. The data are represented by solid black points with statistical uncertainties. The ratios of the data yields after background subtraction to the expected background yields are shown in the lower plots [24]

times branching ratio for dark Higgs boson masses above 160 GeV and this search provides the first CMS experiment results on the dark Higgs model.

In this analysis events with two leptons and large MET are selected. In addition, a set of requirements on event kinematic and topological quantities are imposed to identify events compatible with $s \rightarrow WW$ events. The distributions sensitive to such signal are the transverse mass calculated from the less energetic lepton and the MET, as well as the invariant mass of the dilepton system. These distributions are shown for one SR in Fig. 10.

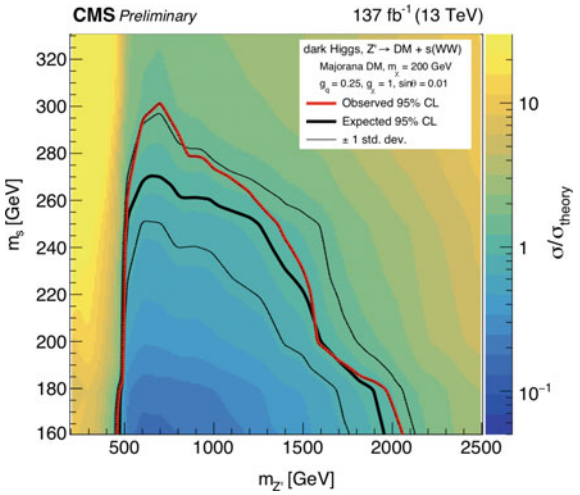

Fig. 11 Upper limits at 95% CL on the model production cross section as a function of the dark Higgs mass m_s (y-axis) and of the Z' mass $m_{Z'}$ (x-axis). The black solid line indicates the observed exclusion boundary for which the ratio μ between the excluded cross section and its theoretical value is equal to 1. All the parameters values enclosed inside this line are excluded by this search [24]

The main backgrounds consist of WW, $t\bar{t}$, and Z(ll) events, which are estimated from data in dedicated CRs. The results are then extracted by a simultaneous fit in SRs and CRs, where the systematic uncertainties are included as nuisance parameters. No significant excess above SM expectation is observed and the results are interpreted in terms of the dark Higgs model introduced above scanning over the dark Higgs, the mediator Z', and the dark matter candidate masses. The corresponding 95% confidence level limits on the production cross section are presented in Fig. 11 for a dark matter mass hypothesis of 200 GeV, where Z' masses $m_{Z'}$ below 2 TeV are excluded.

4.2 Higgs Boson a Portal to the Invisible

Another possibility offered by the Higgs boson, is to study events where the Higgs boson itself is considered as the mediator of the interaction between the dark mater and the Standard Model particles. In this Higgs portal scenario [25–27] the dark matter particles are produced from the decay of the Higgs boson. Higgs portal models are a very attractive theoretical possibility since a decay branching fraction B to invisible particles of up to about 20% is allowed by current constraints [28].

Events where dark matter is produced from the Higgs boson decay will appear as an enhancements of the invisible H decays branching ratio over the Standard Model expectations. Searches for invisible Higgs decays assume Standard Model-like Higgs boson production, where the vector-boson fusion mode has the highest sensitivity. A representative Feynman diagram for vector-boson fusion processes is presented in Fig. 12.

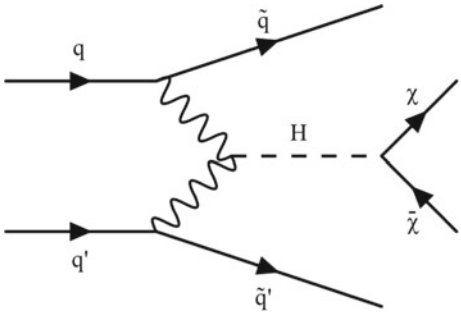

Fig. 12 Representative Feynman diagrams illustrating events where a Standards Model Higgs boson is produced via vector-boson fusion and then it decays into dark matter particles [28]

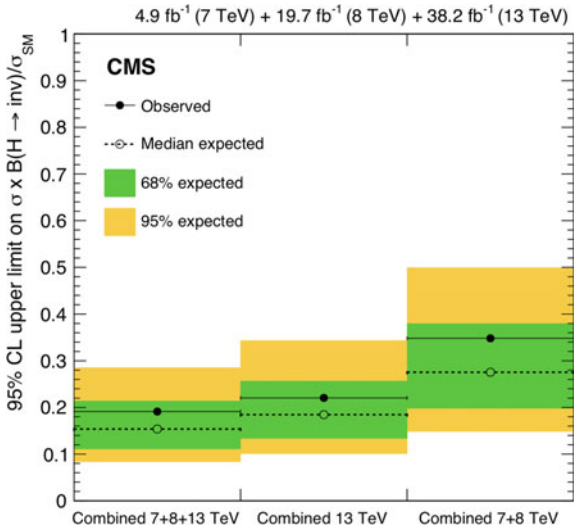

Fig. 13 Upper limits at 95% CL on the ratio between the excluded cross section and its theoretical value times the Higgs to invisible branching ratio. The partial combinations based either on 7+8 or 13 TeV data as well as their combination are presented, assuming Standard Model production cross sections for the Higgs boson [28]

For this type of events [28] high MET and two jets are required. The signal is expected to appear as an excess over the background expectations at large values of the invariant mass of the dijet system. For this reason, it is essential for the experiments to have an excellent calorimetry in the forward region, where the jets are expected. The CMS collaboration has combined the results on H invisible decays from data collected at 7, 8 and 13 TeV center-of-mass energy considering gluon fusion and associated VH production in addition to vector boson fusion processes [28]. The corresponding 95% confidence level upper limits on the ratio between the excluded cross section and its theoretical value times the Higgs to invisible branching ratio are presented in Fig. 13. Higgs to invisible branching ratios above 0.19% are excluded.

The results from the mono-jet search [8], presented above, can also be interpreted under the theoretical assumptions of invisible decays of the Standard Model Higgs boson. The corresponding 95% confidence level upper limits on the ratio between the excluded cross section and its theoretical value times the Higgs to invisible branching ratio are presented in Fig. 14. In the mono-jet category, Higgs to invisible branching ratios larger than 59.6% are excluded, while in the mono-V category values above

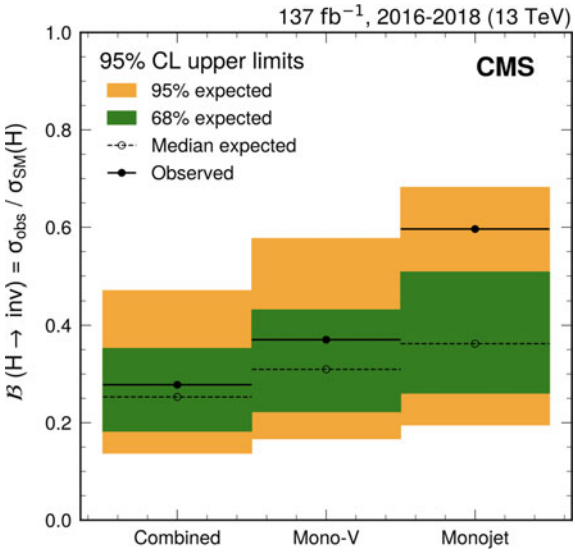

Fig. 14 Upper limits at 95% CL on the ratio between the excluded cross section and its theoretical value times the Higgs to invisible branching ratio. The results are shown separately for the mono-jet and mono-V categories, as well as for their combination [8]

37.0% are excluded. Combining both signatures, Higgs to invisible branching ratios larger than 27.8% are excluded, providing the most stringent to date limit from the combined gluon-fusion and V(qq)H channels.

Acknowledgements The work of D. P. was supported by the U.S. Department of Energy under grant number DE-SC0017647.

References

1. F. Zwicky, Die Rotverschiebung von extragalaktischen Nebeln. Helv. Phys. Acta **6**, 110–127 (1933). https://doi.org/10.1007/s10714-008-0707-4
2. V.C. Rubin, W.K. Ford Jr., Rotation of the andromeda nebula from a spectroscopic survey of emission regions. Astrophys. J. **159**, 379–403 (1970). https://doi.org/10.1086/150317
3. D. Clowe, M. Bradac, A.H. Gonzalez, M. Markevitch, S.W. Randall, C. Jones, D. Zaritsky, A direct empirical proof of the existence of dark matter. Astrophys. J. Lett. **648**, L109–L113 (2006). https://doi.org/10.1086/508162
4. G. Steigman, M.S. Turner, Cosmological constraints on the properties of weakly interacting massive particle. Nucl. Phys. B **253**, 375–386 (1985). https://doi.org/10.1016/0550-3213(85)90537-1
5. S. Chatrchyan et al., The CMS experiment at the CERN LHC. JINST **3**, S08004 (2008). https://doi.org/10.1088/1748-0221/3/08/S08004
6. D. Abercrombie et al., Dark matter benchmark models for early LHC Run-2 searches: report of the ATLAS/CMS dark matter forum. Phys. Dark Univ. **27**, 100371 (2020). https://doi.org/10.1016/j.dark.2019.100371
7. CMS collaboration public results. https://cms-results-search.web.cern.ch
8. A. Tumasyan et al., Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{B} = 13$ TeV. JHEP **11**, 153 (2021). https://doi.org/10.1007/JHEP11(2021)153

9. L. Lopez-Honorez, T. Schwetz, J. Zupan, Higgs portal, fermionic dark matter, and a standard model like Higgs at 125 GeV. Phys. Lett. B **716**, 179–185 (2012). https://doi.org/10.1016/j.physletb.2012.07.017
10. A. Berlin, S. Gori, T. Lin, L.-T. Wang, Pseudoscalar portal dark matter. Phys. Rev. D **92**, 015005 (2015). https://doi.org/10.1103/Phys-RevD.92.015005
11. R.S. Chivukula, H. Georgi, Composite technicolor standard model. Phys. Lett. B **188**, 99–104 (1987). https://doi.org/10.1016/0370-2693(87)90713-1
12. L.J. Hall, L. Randall, Weak scale effective supersymmetry. Phys. Rev. Lett. **65**, 2939–2942 (1990). https://doi.org/10.1103/PhysRevLett.65.2939
13. A.J. Buras, P. Gambino, M. Gorbahn, S. Jager, L. Silvestrini, Universal unitarity triangle and physics beyond the standard model. Phys. Lett. B **500**, 161–167 (2001). https://doi.org/10.1016/S0370-2693(01)00061-2
14. G. D'Ambrosio, G.F. Giudice, G. Isidori, A. Strumia, Minimal flavor violation: an effective field theory approach. Nucl. Phys. B **645**, 155–187 (2002). https://doi.org/10.1016/S0550-3213(02)00836-2
15. A.M. Sirunyan et al., Search for dark matter produced in association with a single top quark or a top quark pair in proton-proton collisions at $\sqrt{B} = 13$ TeV. JHEP **03**, 141 (2019). https://doi.org/10.1007/JHEP03(2019)141
16. A.M. Sirunyan et al., Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques. JINST **15**(06), P06005 (2020). https://doi.org/10.1088/1748-0221/15/06/P06005
17. A.M. Sirunyan et al., Search for resonant and nonresonant new phenomena in high-mass dilepton final states at $\sqrt{B} = 13$ TeV. JHEP **07**, 208 (2021). https://doi.org/10.1007/JHEP07(2021)208
18. A. Tumasyan et al., The CMS collaboration. Search for long-lived particles decaying into muon pairs in proton-proton collisions at $s\sqrt{s} = 13$ TeV collected with a dedicated high-rate data stream. J. High Energ. Phys. **62**, (2022). https://doi.org/10.1007/JHEP04(2022)062
19. D. Curtin, R. Essig, S. Gori, J. Shelton, Illuminating dark photons with high-energy colliders. JHEP **02**, 157 (2015). https://doi.org/10.1007/JHEP02(2015)157
20. S. Chatrchyan et al., Observation of a New Boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B **716**, 30–61 (2012). https://doi.org/10.1016/j.physletb.2012.08.021
21. S. Chatrchyan et al., Observation of a New Boson with mass near 125 GeV in pp collisions at $\sqrt{B} = 7$ and 8 TeV. JHEP **06**, 081 (2013). https://doi.org/10.1007/JHEP06(2013)081
22. A.M. Sirunyan et al., Search for dark matter particles produced in association with a Higgs boson in proton-proton collisions at $\sqrt{B} = 13$ TeV. JHEP **03**, 025 (2020). https://doi.org/10.1007/JHEP03(2020)025
23. M. Duerr et al., How to save the WIMP: global analysis of a dark matter model with two s-channel mediators. JHEP **09**, 042 (2016). https://doi.org/10.1007/JHEP09(2016)042
24. CMS Collaboration, Search for dark matter particles produced in association with a dark Higgs boson decaying into W^+W^- in proton-proton collisions at $\sqrt{S} = 13$ TeV with the CMS detector. CMS-PAS-EXO-20-013. (2021)
25. S. Kanemura, S. Matsumoto, T. Nabeshima, N. Okada, Can WIMP dark matter overcome the nightmare scenario? Phys. Rev. D **82**, 055026 (2010). https://doi.org/10.1103/PhysRevD.82.055026
26. O. Lebedev, H.M. Lee, Y. Mambrini, Vector Higgs-portal dark matter and the invisible Higgs. Phys. Lett. B **707**, 570–576 (2012). https://doi.org/10.1016/j.physletb.2012.01.029
27. A. Djouadi, O. Lebedev, Y. Mambrini, J. Quevillon, Implications of LHC searches for Higgs–portal dark matter. Phys. Lett. B **709**, 65–69 (2012). https://doi.org/10.1016/j.physletb.2012.01.062
28. A.M. Sirunyan et al., Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at $\sqrt{B} = 13$ TeV. Phys. Lett. B **793**, 520–551 (2019). https://doi.org/10.1016/j.physletb.2019.04.025

Probing Stealth Dark Sectors with LHCb

Carlos Vázquez Sierra and José Zurita

1 The Beautifulness of Flavor

To discuss the dark matter search strategies at LHCb, it is useful to first establish the main conceptual design differences between this detector, and the previously described ATLAS and CMS experiments. These differences are due to the distinct physics goals of LHCb.

Sometimes, ATLAS and CMS are referred to as *general-purpose detectors* or *GPDs*. Their design aims at exploring higher energies than the previous generation proton-proton collider (Tevatron, in Fermilab, Batavia, USA) in an agnostic way, collecting as much data as possible, and with the aim of discovering the Higgs boson for masses above 114 GeV (the mark set by the LEP collider in the early 2000s) as well as targeting heavy particles with masses between about 100 GeV and up to a few TeV. The reasons why to expect new particles in these ranges are numerous and they would not fit in the space dedicated to this chapter, but it suffices to say that chartering the unexplored TeV-verse was well-motivated, not only from a dark matter perspective but also due to other theoretical (e.g. electroweak naturalness, matter–antimatter asymmetry, strong *CP*-problem) and experimental (e.g. anomalous magnetic moment of the muon, neutrino masses in the eV range) puzzles.

In contrast, the main physics goal of LHCb is to explore what is known in jargon as *flavor physics*, although in all honesty, it should be called *quark flavor physics*. In the energy range between 2 and 10 GeV approximately, the quarks do not show

C. Vázquez Sierra (✉)
European Organization for Nuclear Research (CERN), Espl. des Particules 1, 1211 Meyrin, Switzerland
e-mail: carlos.vazquez@cern.ch

J. Zurita
Instituto de Física Corpuscular (CSIC-Universitat de Valencia), Catedrático José Beltrán 2, 46980 Valencia, Spain
e-mail: jzurita@ific.uv.es

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_7

up as free particles, but instead, they form bound states, analogous to how protons, neutrons, and electrons appear to us as atoms or molecules. These bound states are held together by the strong force, and the theory describing this force is known as *quantum chromodynamics* or *QCD*.

Due to historical reasons, physicists let their imagination fly and named new quantum numbers based on the human senses, so the quarks have *flavor* (u, d, c, s, t, b) and *color* (red, blue, green), besides other known properties such as electric charge or spin.[1] QCD is a very peculiar force if we try to compare it with the more familiar effects of gravity or electromagnetism. If we consider two masses, the gravitational interaction between them *decreases* if we pull them apart, and *increases* if we move them closer to each other. In QCD instead, the opposite happens: when quarks and gluons are *next to each other* (let us simply say *very close*) they do not exert a force, but when we try to pull them apart there would be a strong force that will put them back together to the same *very close* configuration. This phenomena goes under the name of *asymptotic freedom*, and it is such an important phenomena that their discoverers David J. Gross, H. David Politzer, and Frank Wilczek were awarded the Nobel Prize of Physics in 2004. Hence, when quarks come together at energies in the 2–10 GeV range,[2] the strong force binds them in what we call *hadrons*, that can be classified as *baryons* (bound states of three quarks, the most famous for us being the proton and neutron) or *mesons* (bound states of a quark and antiquark pair, here we find the pions and kaons, for instance). Indeed, there is a large number of such bound states, and flavor physics studies the different transition processes that occur during the *hadron* decays.

The LHCb experiment has been originally designed to study in detail the decays of B mesons (which are bound states of a b quark and an anti-b quark copiously produced at proton-proton collisions) and other hadrons containing b or c quarks, while measuring properties (masses, lifetimes, branching ratios, and other observables) of the different hadrons. To accomplish this goal, LHCb requires an exquisite precision to measure the momenta and lifetimes of these particles at *low energies*. Since the LHC collisions do generate an enormous amount of particles with *low momenta* (see Fig. 1), the LHCb experiment must have a different strategy than that of ATLAS and CMS.

Instead, LHCb (i) is a *single-arm* detector instrumented in the forward region (covering an angle between 10 and 300 mrad), which allows to fully characterize the decay products of B mesons and other heavy flavor hadrons (see Fig. 2), (ii) has the ability to reconstruct decay vertices with a very high spatial resolution (this is, separated by very small distances (of the order of several μm), (iii) is capable of a very precise particle identification that allows to identify individual hadrons such as kaons and pions, and (iv) can measure momenta at a lower threshold than ATLAS and CMS. On top of this, and to allow the detector to cope with such a high number of

[1] Lucky for us, no new quantum numbers have emerged, or we would be speaking of the odor and sound of quarks.

[2] The upper range corresponds to roughly twice the b quark mass. The top quark, having a mass of about 175 GeV, does not hadronize, and we will not be concerned about it in what follows.

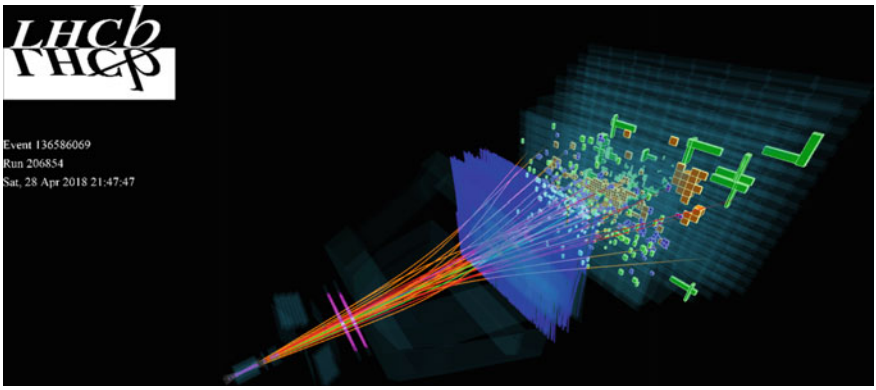

Fig. 1 Typical event display of a heavy flavor process recorded by the LHCb experiment during 2018. The different sub-detector stations (tracking systems, calorimeters, and vertex locator) are shown, as well as the tracks and calorimeter deposits. Different types of particles have different colors associated **(LHCb/CERN)**

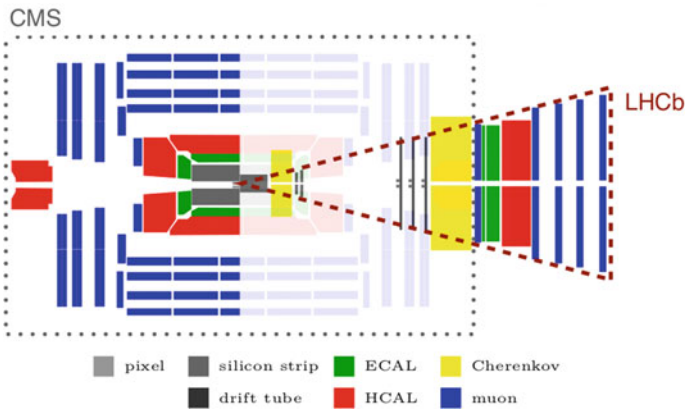

Fig. 2 Schematics of the LHCb and CMS detectors one of top of the other, showing the different coverage regions between the two experiments **(CMS/LHCb/CERN)**

particles at the *low energy regime*, the LHC delivers fewer proton-proton collisions to LHCb than to ATLAS and CMS, leading to a smaller fraction of recorded data (approximately 5–10%) [1].

This *quality data* is optimally selected to fully exploit the flavor physics program but is certainly a price to pay in comparison to ATLAS and CMS when searching for *needle in a haystack* type of rare events.

How can then the BSM program of LHCb rival that from ATLAS and CMS? Well, precisely exploiting the LHCb features that flavor physics requires. For example, needing to identify the flavor of the decay products, having to deal with particles with low momenta, normally arising from the decay of *new neutral particles* with

masses in the 1–100 GeV[3] range or particles that are able to fly a macroscopic appreciable distance before decaying, would become really challenging at ATLAS and CMS (either due to their design or because of the optimization of the reconstruction algorithms). Concrete examples for each of these cases have been collected in the "Unleashing the full power of LHCb to probe Stealth New Physics" Stealth White Paper [2], where the *stealthiness* of the *new physics* is with respect to the GPDs.

In the rest of this chapter, we will then describe a selection of LHCb searches that target several dark matter scenarios that often feature more than one particle, hence we will refer to those as *dark sectors*. These dark sectors would contain one dark matter particle candidate that would be stable, colorless, and electrically neutral, but we will not impose the condition of satisfying the relic density. In other words, these dark sectors correspond to the approach of simplified models detailed in previous introductory chapters. We will hunt for both the dark matter particle itself and also for *mediators*: particles that connect the visible sector (where the Standard Model particles live) with the dark sector, where the dark matter particle resides (in jargon particle physicists also talk about *portals* to the dark sector).

2 On the Hunt for Dark and Light

LHCb has searched for a low mass particle, either *prompt* (decaying within a negligible distance from the proton-proton collision) or *displaced* (flying a certain distance before decaying), and decaying into two muons [3] (see Fig. 3). This low mass particle is studied here in the context of a *dark photon*, a new vector boson (like the photon or the Z-boson) that has a small mixing with the SM photon via a mechanism known as *kinetic mixing* (ε). Leaving technical details aside, this kinetic mixing yields suppressed couplings to the SM particles (due to a myriad of experimental constraints which we display later in Fig. 4), but at the same time, this *dark photon* can act as a portal to the dark sector. The search carried out at LHCb is similar to those performed by ATLAS and CMS (described in previous chapters), which also hunt for this dark photon decaying into a pair of muons of opposite charge, since the dark photon is electrically neutral. In spite of the similar final state, we note that ATLAS and CMS consider instead a different production mechanism (since they study a different dark sector), and hence a one-to-one comparison can not be established. LHCb provides competitive limits in the 1–30 GeV range for the prompt case and world-leading limits for the displaced case down to 214 MeV (di-muon threshold). The latter result could be extended if the analysis considered electrons instead of muons. However, this final state entails additional complications and is being scrutinized by the collaboration.

[3] We will dub this mass interval as *light* or *low mass*, in contrast with ATLAS and CMS that in addition can hunt for particles with masses above 100 GeV and up to few TeV. We stress that both ATLAS and CMS can also hunt for particles in the light regime, but with a few exceptions (muon final states) it is notoriously challenging.

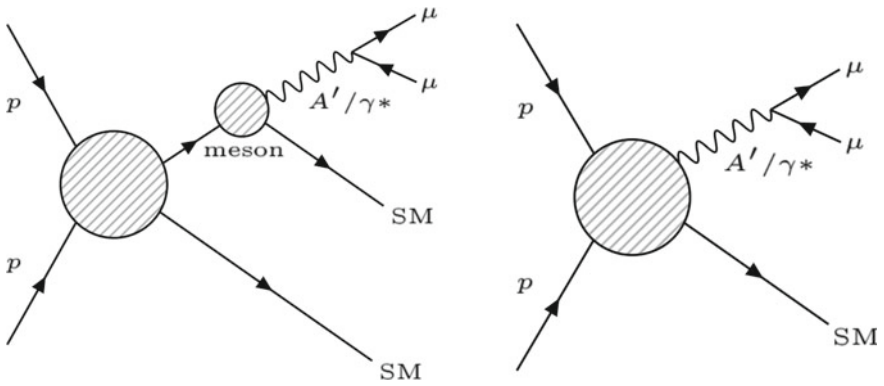

Fig. 3 Production and decay mechanisms of the dark photon (A') described in the text. There are two possible production mechanisms: from a meson decay (left) or directly from proton-proton collisions (right)

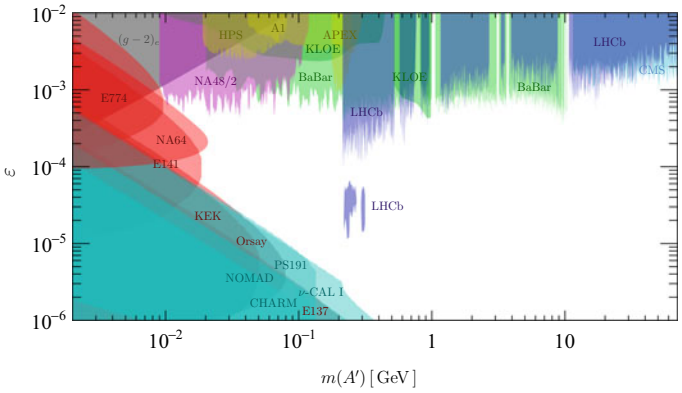

Fig. 4 Comparison of the results to existing constraints from previous experiments **(LHCb/CERN)**

A large dataset collected by the detector from 2016 to 2018 is used, consisting of events containing at least a pair of muons with certain loose requirements on their kinematics. This first step of the selection happens *online*, this is, immediately after data collection from the LHC proton-proton collisions, thanks to the fast LHCb *trigger* system.

Then, pairs of muons are combined to form photon candidates. A more stringent selection consisting of kinematic and particle identification requirements are applied on these candidates, which are also required to form a decay vertex of good quality. This is done to remove candidates that are produced from known processes other than the process of interest, and are known as *background components*. These may consist of muons produced from the decay of heavy-flavor particles, or kaons and pions misidentified as muons by the particle identification system, among others.

One of the most challenging background components in the case of the search for a displaced dark photon are the material interactions with what is known as the radio-frequency foil of the *vertex locator*, or VELO. This envelope, consisting of an aluminum alloy, surrounds the VELO and serves as a physical separation with respect to the LHC beam pipe, suppressing any potential heating or electrical interference on the very delicate silicon sensors of the vertex locator, among others. However, particles from the proton-proton collisions may interact with the foil, leading to the presence of decay vertices that would look like potential candidates of interest. This background component is efficiently suppressed thanks to the use of a very precise mapping of the foil and becomes one of the dominant background components for most of the searches of displaced particles directly produced from proton-proton collisions. In layman terms, we are simply performing an *X-ray* of the LHCb detector!

In the search of *prompt* dark photons, only candidates with masses from 214 MeV to 70 GeV are used. The lower limit corresponds to twice the value of the mass of the muon, which is the lowest possible mass of a particle decaying into a pair of muons. The choice of the upper limit is driven by the fact that the number of events satisfying the selection decreases dramatically after this value, being the available data not enough to perform a precise study. For the case of *displaced* dark photons, an even smaller region of mass candidates is explored, from 214 to 350 MeV, where the upper limit is chosen due to similar efficiency reasons as that for the *prompt* search.

Finally, differences in the selected data with respect to Standard Model predictions are computed, where an excess in these differences would be a hint of the existence of a *new physics* particle. No excess has been found, and a large region of mass and displacement of a hypothetical dark photon has been excluded, as shown in Fig. 4.

Together with this *dark photon* model, other potential scenarios that include the presence of a prompt or displaced particle decaying into two muons have been considered [4]. Scenarios that (a) consider the decay of a *dark hadron* into two muons (where the *dark hadron* is part of a *dark sector*), (b) where the production of this new particle would involve the presence of new Higgs bosons, or (c) where the pair of muons is accompanied by one or two b quarks that hadronize in several final states[4] have been studied, among others. No excess has been found for any of these different scenarios.

3 Dark Sectors, Quarks, and *Jets*

Another interesting signature to look for would consist of a low mass particle decaying into two *quarks*, instead of into two *muons*. The search strategy would be significantly different depending on the quark flavor since *light quarks* (u, d, s) would hadronize in kaons and pions, and *heavy quarks* (c, b) would hadronize in heavy flavor D and B mesons.

[4] Motivated by an existing result [5] from the CMS collaboration.

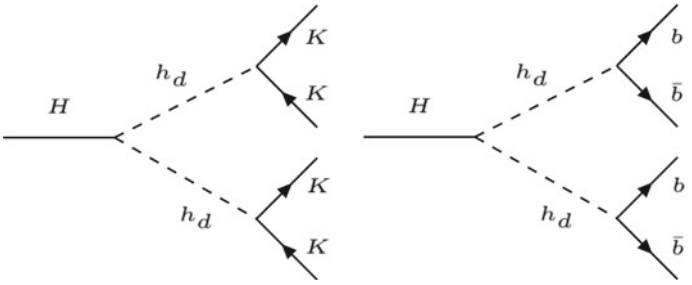

Fig. 5 Schematic of production of two dark hadrons into a pair of quark and an anti-quark, and produced from the decay of a Higgs boson. On the left, the dark hadron decays into light quarks that then hadronize into kaons, while on the right the dark hadron decays into b quarks

Light hadrons such as kaons and pions do not decay within the detector and can be easily reconstructed with high efficiency as *one track per hadron* thanks to the tracking and particle identification systems. A potential search involving a *displaced* low mass particle decaying into pairs of kaons has been proposed [6]. Here, this new particle would be produced in pairs from the decay of a Higgs boson, and then decay into two kaons (see Fig. 5). Using only simulated data, a suitable strategy, as well as projections on the potential exclusion of this search, are studied, showing the capabilities of LHCb to provide competitive results for this kind of signature. Although still not complete, the collaboration is preparing an analysis in this regard towards a publication in the near future.

On the other hand, B and D mesons are *unstable* particles that decay within the detector, either into other unstable particles (that also decay) and/or into stable particles. This leads to a situation where a large number of decay vertices have to be reconstructed to characterize a single heavy flavor quark, having an impact on the reconstruction efficiency. To achieve a reasonable reconstruction efficiency, the so-called *jet reconstruction technique* is used instead, making use of all the particle tracks produced from the decay vertex of the quark without having to reconstruct all the decay vertices.

LHCb has searched for a hypothetical *dark hadron* decaying into *two heavy flavor quarks,* produced in pairs from the decay of a Higgs boson [7], as sketched in Fig. 5. First, the dataset collected between 2011 and 2012 is filtered by requiring events to have a *displaced vertex* with a minimum number of associated tracks. This helps to remove most of the *background contributions*, including those from *material interactions* as previously described. Then, the jet reconstruction technique is used to reconstruct jets from heavy-flavor quarks, and pairs of jets are combined to obtain the dark hadron candidates. Only *displaced* candidates with masses between 25 and 50 GeV are considered, due to efficiency reasons. A similar but not the same technique to that of the search for a *dark photon* is finally used, in order to search for any potential excess: a complementary region of masses and lifetimes to those studied by ATLAS and CMS is studied, and no excess has been found as shown in Fig. 6.

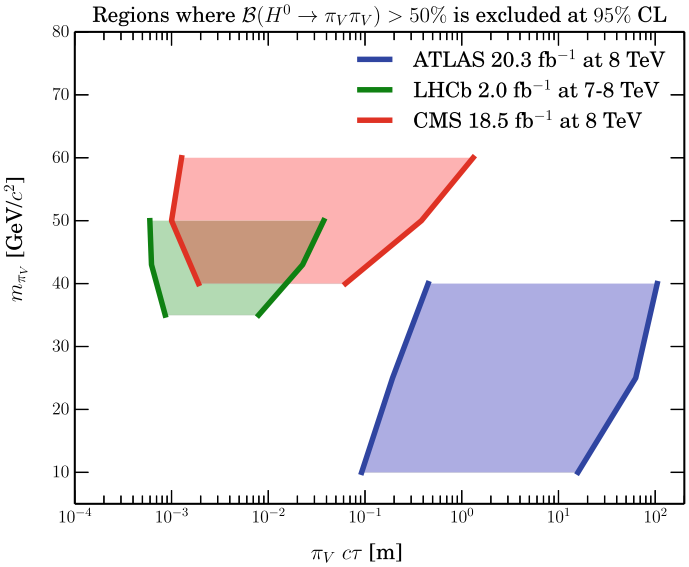

Fig. 6 Comparison of the results to some of the existing constraints from other experiments **(LHCb/CERN)**

4 Matter, Antimatter and Dark Matter

Finally, a very interesting model that attempts to both explain the matter–antimatter asymmetry in the Universe and the existence of dark matter, has been recently proposed [8]. The mechanism behind this model, known as *B-mesogenesis*, postulates a dark matter candidate produced from the decay of a B meson (among other particles containing a b quark), together with other ordinary particles detectable at a collider experiment (see Fig. 7). Because of its unique design (due to the physics

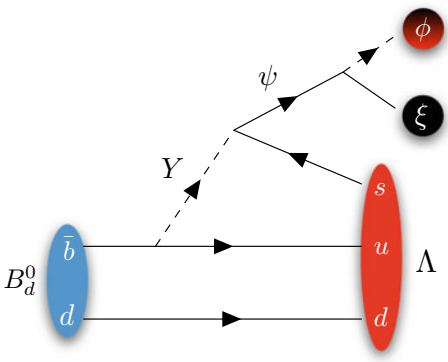

Fig. 7 Example of the decay of a B meson into a Λ baryon (detectable), and a dark matter candidate ψ which remains invisible to the detector [8]

goals of studying flavor), the LHCb detector becomes a natural candidate to search for this kind of signature.

The capabilities of LHCb to test this model have been discussed in [9], where a search strategy was proposed and potential exclusion limits were obtained using simulated data. One of the main differences of this search with respect to those previously described in this section is the fact that the hypothetical *dark matter* particle, invisible and hence not reconstructable, would be produced from the decay of *another unstable particle*, and not directly from proton-proton collisions: constraints on the displacement of the mass and decay vertex of the parent particle (a B meson) are extremely helpful and can be exploited as part of the selection strategy. Then, the rest of the strategy would consist of reconstructing the accompanying products of the decay of the B meson. With all these ingredients and in a similar way as ATLAS and CMS do, an excess on the distribution of *missing energy* can be measured, which would correspond to that of the invisible dark matter particle.

The exclusion potential from this study shows very encouraging prospects towards a future search at LHCb, which is already ongoing and to be completed in the near future.

5 Contributing to the Hunt for Dark Matter

In brief, the LHCb detector, originally conceived to explore the intricate hadronic world, can exploit its design capabilities to hunt for dark sectors with (a) particles in the 1–100 GeV range decaying either prompt or long-lived, (b) decay products with low transverse momentum and (c) dark particles stemming directly from B-meson decays. These are just a sample of an expanding physics program at LHCb and one can expect more searches in the near future, probing previously uncharted dark lands.

References

1. F. Follin, D. Jacquet, Implementation and experience with luminosity levelling with offset beam, in *CERN Yellow Report CERN-2014-004* (2014), pp.183–187. https://doi.org/10.5170/CERN-2014-004.183
2. M. Borsato, X.C. Vidal, Y. Tsai et al., Unleashing the full power of LHCb to probe stealth new physics. Rep. Prog. Phys. 85 024201 (2022). https://doi.org/10.1088/1361-6633/ac4649, https://arxiv.org/abs/2105.12668
3. LHCb Collaboration, Search for $A' \to \mu^+\mu^-$ decays. Phys. Rev. Lett. **124**, 041801 (2020). https://doi.org/10.1103/PhysRevLett.124.041801
4. LHCb Collaboration, Searches for low-mass dimuon resonances. J. High Energ. Phys. **156** (2020). https://doi.org/10.1007/JHEP10(2020)156
5. The CMS Collaboration, Search for resonances in the mass spectrum of muon pairs produced in association with b quark jets in proton-proton collisions at $\sqrt{s}$= 8 and 13 TeV. J. High Energ. Phys. **161** (2018). https://doi.org/10.1007/JHEP11(2018)161

6. X.C. Vidal, Y. Tsai, J. Zurita, Identifying exclusive displaced hadronic signatures in the forward region of the LHC. J. High Energ. Phys. **2020**, 115 (2020). https://doi.org/10.1007/JHEP01(2020)115
7. LHCb Collaboration, Updated search for long-lived particles decaying to jet pairs. Eur. Phys. J. C **77**, 812 (2017). https://doi.org/10.1140/epjc/s10052-017-5178-x
8. G. Elor, M. Escudero, A.E. Nelson, Baryogenesis and dark matter from B mesons. Phys. Rev. D **99**, 035031 (2017). https://doi.org/10.1103/PhysRevD.99.035031
9. A. Brea Rodríguez, V. Chobanova, X.C. Vidal et al., (2016) Prospects on searches for baryonic Dark Matter produced in b-hadron decays at LHCb. Eur. Phys. J. C **81**, 964 (2021). https://doi.org/10.1140/epjc/s10052-021-09762-w

Hunting Dark Matter Axions with CAST

Marios Maroudas and Kaan Ozbozduman

1 Introduction

One of the most fundamental problems of Modern Physics is that of the composition of Dark Matter (DM) which accounts for approximately 85% of the total matter content in the universe. One of the most promising particle candidates for DM are the axions. They are hypothetical elementary particles that were initially postulated in 1978 resulting from the Peccei–Quinn mechanism [1] which was introduced as a solution to the strong CP problem of the Standard Model. Axions, which got their name after a detergent, if they exist they must have a very small mass and must interact very feebly with normal matter. At the same time, they can sufficiently be produced during the Big Bang making them ideal candidates for cold DM.

CERN's Axion Solar Telescope (CAST) started searching for axions coming from our Sun in 2003. The detection principle is based on the inverse Primakoff effect, where, in the presence of a strong magnetic or electric field, the axion can convert into a photon. Cutting-edge limits on the axion-photon coupling were then set by CAST during its operation. As seen in Fig. 1, the latest upper limit on the axion-photon coupling for axion masses bellow 0.02 eV is $0.66 \times 10^{-10}\,\mathrm{GeV}^{-1}$ [2].

In 2019, following a suggestion from 2012 [3] CAST was transformed from an axion helioscope looking for solar axions to an axion haloscope looking for DM axions in the μeV mass region. This was based on the Sikivie haloscope technique [4], where, in the presence of a strong magnetic field, axions from the galactic halo convert into photons if the resulting photons are detected inside a high-quality

M. Maroudas (✉)
Department of Physics, University of Patras, 26504 Patras, Greece
e-mail: marios.maroudas@cern.ch

K. Ozbozduman
Institute of Sciences, Istinye University, 34396 Sariyer, Istanbul, Turkey
e-mail: kaan.ozbozduman@cern.ch

Physics Department, Bogazici University, 34342 Bebek, Istanbul, Turkey

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_8

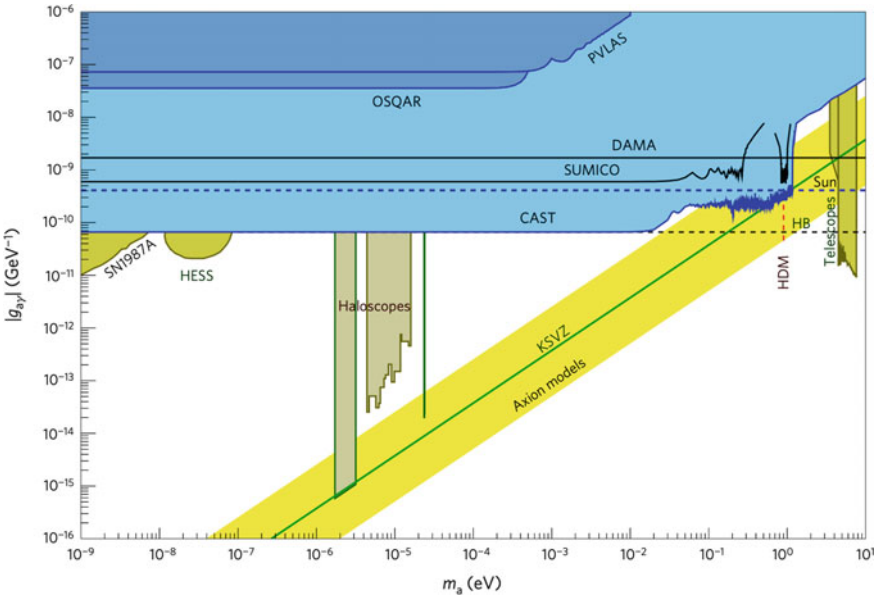

Fig. 1 The solar axion exclusion plot set by CAST while pointing at the Sun [2]

microwave cavity resonating to the corresponding frequency defined by the unknown axion rest mass. The microwave signal is then extracted through an antenna which is critically coupled to the cavity. Since the axion mass is unknown, haloscopes should be tuneable in order to be able to change the cavity's resonant frequency and thus scan a wide range of axion masses.

The CAST superconducting dipole magnet provides the strong external magnetic field of 8.8 T and has a twin-bore geometry into which rectangular cavities are fitted. The probability of a DM axion to be converted into a real photon inside a microwave cavity, increases with the square of the magnetic field (B^2), the quality factor of the cavity (Q) which is the ratio of the cavity stored energy to its power loss per cycle, the volume of the cavity (V) and the geometry factor (C) which is determined by the direction of the external magnetic field and the cavity mode used:

$$P_{\text{axion}} \approx g_{a\gamma\gamma}^2 \frac{\rho_a}{m_a} B^2 Q V C \tag{1}$$

where ρ_a is the local mass density of DM axions, m_a the mass of the axion and $g_{a\gamma\gamma}$ the axion-photon coupling.

The conventional search for DM axions has been so far based on the assumed isotropic halo distribution of our galaxy with the local DM density ρ_a usually assumed to have an average value of 0.45 GeV/cm^3 [5, 6]. However, this could be the reason why axions have not been detected so far. As we will see also in Sect. 2, considering axion DM streams [7] propagating near the ecliptic plane of our solar system and

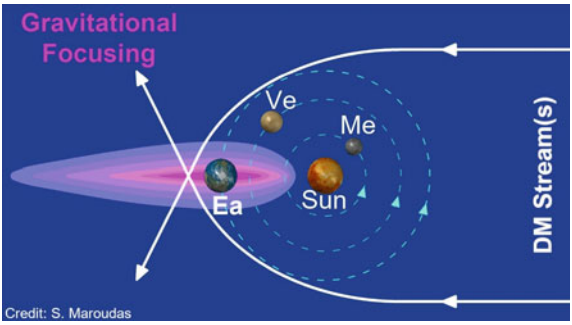

Fig. 2 Schematic view of the flow of a slow-moving DM stream being gravitationally focused by the Sun towards the Earth resulting in flux enhancements of several orders of magnitude (© Marios Maroudas 2022. All rights reserved)

towards an Earth bound DM axion detector like CAST-CAPP, large flux enhancements can take place due to gravitational focusing effects by the solar system bodies including the Sun itself (see Fig. 2) [8, 9].

To take advantage of such burst-like axion flux enhancements due to temporally occurring stream alignments with the Earth, two criteria must be fulfilled:

1. The covered frequency range must be as wide as possible since the axion mass is unknown.
2. The scanning time must be as short as possible in order to take advantage also of short-lasting alignments towards the Earth between the stream and the intervening solar system body.

2 Methodology

In 2019 two different microwave cavity detectors were installed inside each one of the two bores of CAST's dipole magnet, CAST-RADES and CAST-CAPP (see Fig. 3), making CAST the only experiment at CERN searching directly for DM. The CAST-RADES sub-detector consists of a 1 m long cavity comprised of alternating irises searching for DM axions around 34.67 μeV [10]. On the other hand, the CAST-CAPP sub-detector, on which we focus here, consists of four rectangular stainless steel cavities, each with a volume of 224 cm^3, and the ability to be tuned in a quite wide range of axion rest mass of about 660 MHz.

As shown in Fig. 4, CAST-CAPP is a unique axion detector containing a delicate fast-tuning mechanism inside each cavity, consisting of two parallel sapphire strips which are displaced by a piezoelectric motor through a locomotive mechanism providing a tuning resolution of less than 100 Hz in stable conditions. The sapphire strips are symmetrically placed parallel to the cavity longitudinal sides moving simultaneously towards the centre. The maximum scanning speed reached with CAST-CAPP is 10 MHz/min with the coverage of the full frequency range taking about 1 h. Thus, if the axion rest mass is within this range, CAST-CAPP can search also for streaming

Fig. 3 The CAST experiment with a close up photo of the twin bores where CAST-RADES and CAST-CAPP microwave cavities are installed (CAST credits)

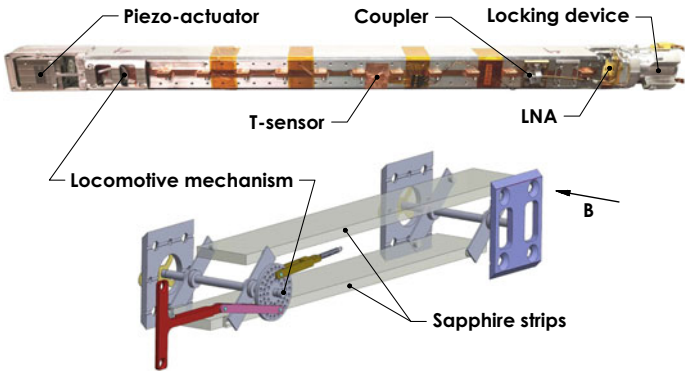

Fig. 4 The CAST-CAPP cavity assembly (top) and its tuning mechanism with the two sapphire strips (bottom) (CAST credits)

DM axions with enhanced sensitivity due to higher axion densities by up to several orders of magnitude.

Smaller cavities are required in order to reach higher axion masses. However, as follows from Eq. 1, the detection sensitivity increases with the cavity volume. To mitigate this issue, CAST-CAPP is using four identical cavities together with the phase-matching technique to increase the effective volume. This technique, which is introduced for the very first time in axion research, improves linearly the signal-to-noise ratio with the number of cavities [11]. To achieve this, a coherent combination of the simultaneous power outputs from the four frequency-matched cavities has to be performed in data-taking conditions.

Using these two novel techniques of fast-frequency tuning and phase-matching, during searches for conventional DM axions, CAST-CAPP became sensitive also to transient events such as axion streams [7] and cosmologically motivated axion mini-clusters [8, 12, 13]. These can give rise to temporally enhanced flux densities (ρ_a)

by several orders of magnitude, in particular when combined with the gravitational lensing effects by the solar system as seen in Fig. 2 [14, 15]. In the ideal case, the flux enhancement due to the gravitational focusing of the Sun can be as high as $\sim 10^{11}$ whereas from the intrinsic Earth mass distribution the enhancement can be up to $\sim 10^{9}$ [16]. Apparently, the faster the scanning the shorter the axion bursts that can be utilized, making the fast-tuning mechanism of CAST-CAPP an indispensable component.

3 Results

From 09/2019 to 06/2021 CAST-CAPP has taken about 172 d of data with both single and phase-matched cavities in data taking conditions with $B = 8.8\,\mathrm{T}$. The scanned frequency range extended from 4.77 to 5.43 GHz covering a parameter phase space of ~660 MHz. This corresponds to axion masses between 19.74 and 22.47 μeV. At the same time background data were taken with $B = 0\,\mathrm{T}$ for about 16 d to exclude possible axion candidates.

Several quality checks were also applied in these data to ensure that undesired effects such as mechanical vibrations are removed from the from further consideration in the analysis. The applied criteria resulted in rejecting about 4.4% of the recorded data.

The performed data analysis was based on widely-accepted methods [6, 17, 18], but was adjusted to the specific experimental conditions of CAST-CAPP. The derived

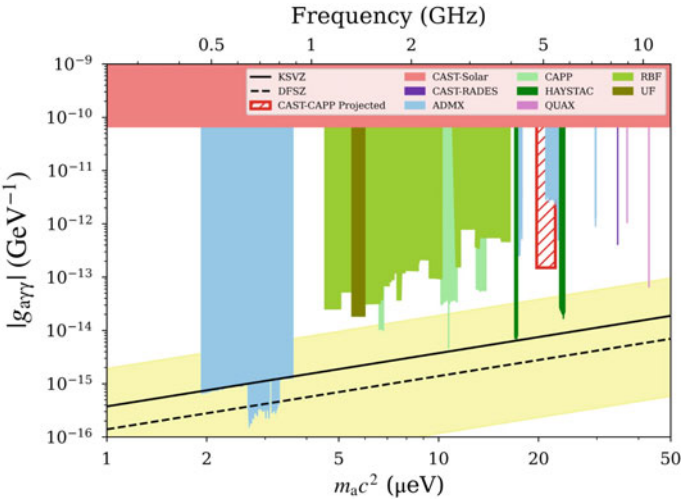

Fig. 5 CAST-CAPP projected exclusion limit on the axion-photon coupling as a function of axion mass compared to other axion search results [2, 17–27]

results showed no galactic DM axion candidate signal above the predefined 5σ level. Therefore, as seen in Fig. 5 new limits on the axion-photon coupling as a function of the axion mass were set, with the achieved performance being competitive with other state-of-the-art DM axion detectors. These results have been submitted for publication and are under consideration.

At the same time, an independent analysis is pending for the next few months, which will allow to search for transients due to aforementioned axion streams or mini clusters. The preliminary results also showed no significant axion lines. However, this novel analysis procedure has to be optimized and therefore it could give a surprise!

4 Conclusions

CAST has been searching for axions for about 22 years. It has progressively set stronger and stronger limits on the solar axion interaction strength, becoming a point of reference in axion research. CAST never stopped evolving throughout the years upgrading its instrumentation and improving its sensitivity for axions, axion-like particles and chameleons [2, 28–31], becoming recently also a DM axion antenna using microwave cavities. The recent competitive results of CAST-RADES [19] and CAST-CAPP were able to set world-class limits on the galactic DM axion-photon conversion and laid the foundations for a search of short-lasting transient events by making use of the two newly developed techniques of fast scanning and phase-matching.

Even though the axion has still eluded CAST efforts, the huge experience gained over the years together with the new technologies and experimental approaches that were introduced will help define the future axion searches with next-generation helioscopes and haloscopes.

Acknowledgements For M.M part of this research is co-financed by Greece and the European Union (European Social Fund—ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning" in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research—2nd Cycle" (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

References

1. R.D. Peccei, H.R. Quinn, Constraints imposed by CP conservation in the presence of pseudoparticles. Phys. Rev. **16**, 1791–1797 (1977). https://doi.org/10.1103/PhysRevD.16.1791
2. V. Anastassopoulos et al., (CAST Collaboration), New CAST limit on the axion-photon interaction. Nat. Phys. **13**, 584–590 (2017). https://doi.org/10.1038/nphys4109
3. O.K. Baker et al., Prospects for searching axionlike particle dark matter with dipole, toroidal, and wiggler magnets. Phys. Rev. D **85**, 035018 (2012). https://doi.org/10.1103/PhysRevD.85.035018

4. P. Sikivie, Experimental tests of the "invisible" axion. Phys. Rev. Lett. **51**, 1415–1417 (1983). https://doi.org/10.1103/PhysRevLett.51.1415
5. E.I. Gates, G. Gyuk, M.S. Turner, The Local halo density. Astrophys. J. Lett. **449**, L123–L126 (1995). arXiv:astro-ph/9505039. https://doi.org/10.1086/309652
6. S. Asztalos et al., (ADMX Collaboration), Large-scale microwave cavity search for dark-matter axions. Phys. Rev. D **64**, 092003 (2001). https://doi.org/10.1103/PhysRevD.64.092003
7. M. Vogelsberger, S.D.M. White, Streams and caustics: the fine-grained structure of Λ cold dark matter haloes. Mon. Not. R. Astron. Soc. **413**(2), 1419–1438 (2011). https://doi.org/10.1111/j.1365-2966.2011.18224.x
8. K. Zioutas et al., Search for axions in streaming dark matter. arXiv:1703.01436
9. H. Fischer, Y. Semertzidis, K. Zioutas, Search for axions in streaming dark matter, CERN EP Newsletter
10. A. Álvarez Melcón et al., Scalable haloscopes for axion dark matter detection in the 30μeV range with RADES. JHEP **07**, 084 (2020). arXiv:2002.07639. https://doi.org/10.1007/JHEP07(2020)084
11. J. Jeong et al., Phase-matching of multiple-cavity detectors for dark matter axion search. Astropart. Phys. **97**, 33–37 (2018). https://doi.org/10.1016/j.astropartphys.2017.10.012
12. I. Tkachev, On the possibility of bose-star formation. Phys. Lett. B **261**(3), 289–293 (1991). https://doi.org/10.1016/0370-2693(91)90330-S
13. E.W. Kolb, I.I. Tkachev, Axion miniclusters and bose stars. Phys. Rev. Lett. **71**, 3051–3054 (1993). https://doi.org/10.1103/PhysRevLett.71.3051
14. D. Hoffmann, J. Jacoby, K. Zioutas, Gravitational lensing by the sun of non-relativistic penetrating particles. Astropart. Phys. **20**(1), 73–78 (2003). https://doi.org/10.1016/S0927-6505(03)00138-5
15. B.R. Patla et al., Flux enhancement of slow-moving particles by sun or Jupiter: can they be detected on earth? Astrophys. J. **780**(2), 158 (2013). https://doi.org/10.1088/0004-637x/780/2/158
16. Y. Sofue, Gravitational focusing of low-velocity dark matter on the Earth's surface. Galaxies **8**(2), 42 (2020). arXiv:2005.08252. https://doi.org/10.3390/galaxies8020042
17. B.M. Brubaker et al., (HAYSTAC Collaboration), Haystac axion search analysis procedure. Phys. Rev. D **96**, 123008 (2017). https://doi.org/10.1103/PhysRevD.96.123008
18. C. Bartram et al., (ADMX Collaboration), Axion dark matter experiment: Run 1b analysis details. Phys. Rev. D **103**, 032002 (2021). https://doi.org/10.1103/PhysRevD.103.032002
19. A. Álvarez Melcón, et al., (CAST Collaboration), First results of the cast-rades haloscope search for axions at 34.67 μev. J. High Energy Phys. (2021). https://doi.org/10.1007/JHEP10(2021)075
20. C. Bartram, et al., (ADMX Collaboration), Dark matter axion search using a josephson traveling wave parametric amplifier. arXiv:2110.10262
21. C. Boutan et al., (ADMX Collaboration), Piezoelectrically tuned multimode cavity search for axion dark matter. Phys. Rev. Lett. **121**, 261302 (2018). https://doi.org/10.1103/PhysRevLett.121.261302
22. K.M. Backes et al., A quantum enhanced search for dark matter axions. Nature **590**(7845), 238–242 (2021). https://doi.org/10.1038/s41586-021-03226-7
23. S. Lee et al., Axion dark matter search around 6.7μeV. Phys. Rev. Lett. **124**, 101802 (2020). https://doi.org/10.1103/PhysRevLett.124.101802
24. J. Jeong et al., Search for invisible axion dark matter with a multiple-cell haloscope. Phys. Rev. Lett. **125**, 221302 (2020). https://doi.org/10.1103/PhysRevLett.125.221302
25. O. Kwon et al., First results from an axion haloscope at CAPP around 10.7μeV. Phys. Rev. Lett. **126**, 191802 (2021). https://doi.org/10.1103/PhysRevLett.126.191802
26. S. De Panfilis et al., Limits on the abundance and coupling of cosmic axions at $4.5<m_a<5.0$ μeV. Phys. Rev. Lett. **59**, 839–842 (1987). https://doi.org/10.1103/PhysRevLett.59.839
27. C. Hagmann et al., Results from a search for cosmic axions. Phys. Rev. D **42**, 1297–1300 (1990). https://doi.org/10.1103/PhysRevD.42.1297

28. G. Cantatore et al., (CAST Collaboration), Search for solar axion like particles in the low energy range at CAST. Nucl. Instrum. Meth. A **617**, 502–504 (2010). https://doi.org/10.1016/j.nima.2009.10.098
29. V. Anastassopoulos et al., (CAST Collaboration), Search for chameleons with CAST. Phys. Lett. B **749**, 172–180 (2015). https://doi.org/10.1016/j.physletb.2015.07.049
30. V. Anastassopoulos et al., (CAST Collaboration), Improved search for solar chameleons with a GridPix detector at CAST. J. Cosmol. Astropart. Phys. JCAP01, (2019)032 https://doi.org/10.1088/1475-7516/2019/01/032
31. S. Arguedas Cuendis et al., (CAST Collaboration), First results on the search for chameleons with the KWISP detector at CAST. Phys. Dark Univ. (2019)100367 https://doi.org/10.1016/j.dark.2019.100367

FASER: The Lifetime Frontier at the LHC and the Search for Dark Matter

Michaela Queitsch-Maitland

Searches for new particles at the large general-purpose experiments at the LHC typically target relatively heavy, strongly interacting particles. If the hypothetical new particles related to dark matter are instead light and weakly interacting, they could completely evade detection in these searches. Such particles would be even more challenging to detect than Weakly Interactive Massive Particles (WIMPs) currently searched for at the LHC and elsewhere. Another possibility is that other particles, known as portal particles, exist that would be the bridge between normal Standard Model (SM) particles and light dark matter. These portal particles could communicate with both SM and dark matter particles, and could be easier to directly detect than dark matter. The probability for producing and detecting these particles is, however, still extremely small and so extraordinarily high event rates would be needed to discover them. In 2017 a team of theorists—Feng, Galon, Kling, and Trojanowski—realised something very exciting: that these portal particles might already be being produced in significant quantities in decays of light particles, such as kaons or pions, at the LHC.

Heavy, strongly interacting new particles are expected to be produced at large angles with respect to the beam direction in proton-proton collisions. The large LHC detectors were designed and built to surround the collision points to ensure the best chance of detecting these particles. These detectors have an important blind spot though—directly down the beam pipes. In collisions at the LHC, the majority of light hadrons like kaons and pions are produced at a small angle relative to the beam direction. It could therefore be that the portal particles produced in pion decays were completely escaping detection by travelling down the beam pipe. The LHC is a ring, with magnets directing the protons around in a circle again and again in order to collide at the main experiments. As portal particles are electrically neutral they

M. Queitsch-Maitland (✉)
CERN, Meyrin, Switzerland
e-mail: michaela.queitsch-maitland@cern.ch

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_9

would be unaffected by the bending magnets of the LHC and continue in a straight line while the beam line curves away. These long-lived, light particles interact so rarely that they can travel hundreds of metres from their production point before interacting and decaying. However, at a distance of 500 m from the collisions where the portal particles are produced, these escaping particles would only have spread out only around 10 cm from one another. A small, well-placed detector would therefore be able to capture and detect the decays of these portal particles.

The perfect location for such a detector was soon located by Feng and his team: a tunnel known as TI12, which used to supply electrons and positrons from the CERN Super Proton Sychrotron (SPS) to the Large Electron-Positron Collider (LEP). The LEP collider operated from 1989 to 2000 and occupied the tunnel now used by the LHC. When building the LHC new injector lines were installed between the SPS and the LHC so the TI12 tunnel was sealed and no longer used. The TI12 tunnel is located around 480 m from the ATLAS interaction point (IP), directly in the straight-line path that the portal particles would take from the collision point (Fig. 1). This was an exciting discovery, and Feng and his team proposed building a new detector in this ideal location [1]. In 2019, the experiment was approved by the CERN experiment board and funding was secured from the Heising and Heising-Simons foundations. The newest of the eight particle physics experiments at CERN's Large Hadron Collider, the Forward Search Experiment (FASER) was born [2].

The next challenge was to build, test, and install the detector. Due to the location of TI12, it was not possible to do this while the LHC machine was actively running. There was also civil engineering work required where the detector would be placed, and other infrastructure to install. The LHC routinely turns off in order to perform maintenance and upgrades. It had shut down at the end of 2018 and was due to start again in early 2021, so the newly formed FASER Collaboration had a tight and very hard deadline by which to get the experiment up and running. As FASER is a small experiment the LHC had to match the LHC beam schedule – not the other way around! This timeline was an important factor in the design of the detector, so FASER was built using existing detector components where possible. When the large LHC experiments were built a number of back-up parts were made, so that in case any parts of the detector developed faults these could be replaced with the spare components. Some surplus back-ups from the ATLAS and LHCb experiments were generously donated to FASER and used to construct the silicon charged particle tracking detector and the calorimeter used to make energy measurements. Using these existing components allowed the FASER detector to be built on schedule. After some delays to the FASER and LHC schedules due to the COVID-19 pandemic, the detector was successfully installed and tested in the tunnel in April 2021. The LHC turns on again in 2022 and FASER will then start recording first collision data.

FASER's main purpose is to detect dark matter portal particles, like dark photons, produced by the LHC, but it will also be able to provide insights into other types of particles [3]. It could find hypothetical particles called axions. Light versions of axions are dark matter candidates, but the axions that FASER will be able to detector are likely too heavy and unstable to be dark matter. It will also be able to detect neutrinos. Although neutrinos are typically considered to be too light to be ideal

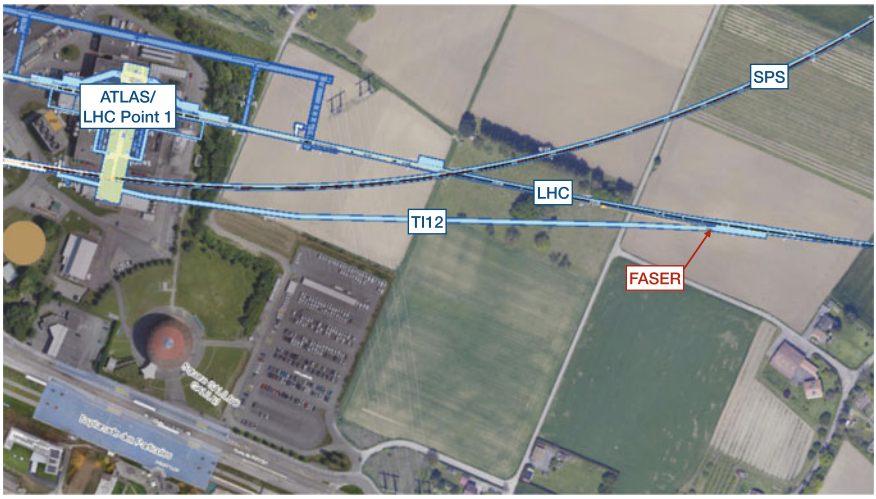

Fig. 1 Aerial view of the CERN campus with the underground experiment and LHC structures superimposed. The location of the FASER experiment relative to the Esplanade des Particules and the ATLAS experiment LHC interaction point 1 (IP1) is shown

dark matter candidates, FASER will be able to provide important insights into the interactions of neutrinos in an energy range that are inaccessible to other experiments using a dedicated neutrino detector FASERν. The FASER detector may also be able to detect decays of hypothetical heavy neutrinos, called sterile neutrinos. Sterile neutrinos are believed to only interact with gravity, and in many theories are either dark matter candidates or act as portal particles to a dark sector containing dark matter. There have been many searches for sterile neutrinos across a wide range of masses and interaction strengths from dedicated neutrino experiments such as MiniBooNE, IceCube, MINOS, to collider experiments to observations of supernovae. So far none of these have observed sterile neutrinos, but FASER will be able to look for them in a mass range that has not been possible to search before.

1 The FASER Detector

FASER will detect hypothetical new long-lived particles, such as dark photons, through their decays to known particles. Dark photons would be produced in decays of hadrons, such as pions, in proton-proton collisions in ATLAS or alternatively through particles hitting the neutral particle absorber located 140 m away from the ATLAS IP. They would then travel directly down the beam pipe, in a tangent to the LHC ring, for 480 m and eventually through 10 m of concrete and 90 m of rock to the FASER detector. The portal particle would then decay within the detector to particles, such as electrons, which could be detected.

For a given particle, its mass and interaction strength determine the probability that it will decay after travelling a given distance, related to the lifetime of the particle. The location, size, and design of the detector are therefore some of the driving factors in the sensitivity of FASER. The signature of the signal is the detection of two highly energetic, oppositely-charged particles (or two neutral photons) that originate from a common production point inside the detector. Using the principle of conservation of momentum, the origin of the parent particle can also be calculated. This should point back to the ATLAS IP where it would originate from. The detector was designed to be able to measure the trajectories and energies of these decay particles.

At the LHC, protons circulate in the machine in bunches and are brought together at the collision points. The time at which a bunch of proton beams is crossed with another—known as a bunch crossing—is recorded by a precise clock. The FASER detector can use this LHC clock to determine when a collision happened and therefore if a signal event was expected. If a signal is recorded in FASER but no collision happened in ATLAS then this was likely a background process. If a signal is recorded in time with a collision in ATLAS then this could be a signal event.

A schematic of the detector is shown in Fig. 2. The detector is approximately 5 metres long and 20 cm in diameter. At the entrance to the detector, two veto scintillator stations (shown in yellow) can be used to detect and veto any charged particles coming from the IP. A scintillator is a material that converts high energy radiation to visible light. A photomultiplier tube (PMT) then converts the outcoming light into an electrical signal. The FASER scintillators have been measured to be over 99.99% efficient and are expected to be mostly used to veto background events containing muons produced in collisions in ATLAS.

Also in front of the main FASER detector is the FASER neutrino emulsion detector, FASERν (shown in pink), and interface tracking station (shown in orange) [4]. This detector contains interlaced 25 cm square layers of emulsion films and 1 mm thick tungsten plates and weighs 1.2 tons. The tungsten plates induce neutrino inter-

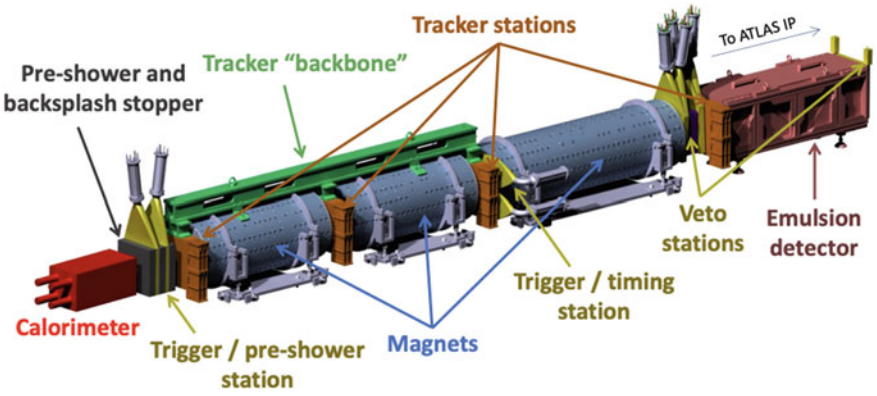

Fig. 2 Labelled diagram of the FASER detector showing the neutrino emulsion detector (FASERν), the 0.55 Telsa magnets, scintillators, silicon tracking stations, and the calorimeter

actions and the emulsion layers records the trajectories of the charged particles produced by the neutrinos. Tungsten is used due to its extremely high density, which increases the chance for interactions while allowing to keep the detector small. The emulsion films are like extremely precise photographic film, comprising of silver bromide crystals with diameters of 200 nm dispersed in gelatin media. Each crystal works as an independent detection channel and the expected resolution of the detector is about 0.4 microns. As the emulsion films will record the trajectories of all particles that pass through it, they will be replaced approximately every three months. The emulsion films will then be digitised and the charged particle trajectories reconstructed. The interface tracking station is the same design as the other tracking stations in the main FASER spectrometer and is used to match charged particle trajectories between FASERν and the rest of the FASER detector.

Next is a 1.5 m long, 0.55 Tesla permanent dipole magnet (shown in blue) with a hollow centre of radius 10 cm. This is where the portal particles are expected to decay into a pair of charged particles. The magnet exerts a bending force on the charged particles, causing them to separate and make them easier to identify in the subsequent parts of the detector.

After the decay volume is a spectrometer. This consists of two 1 m long 0.55 Tesla magnets (shown in blue) and three tracking stations (shown in orange). Each tracking station is comprised of precision silicon micro-strip detectors [5]. The silicon detectors are used to measure the position of charged particles. Combined with the magnetic fields the FASER spectrometer can also measure the momentum of charged particles passing through the detector. Each station is comprised of three layers of eight silicon micro-strip modules donated by the ATLAS SCT Collaboration. The three tracking stations are supported on a mechanical backbone (shown in green), which is mounted on the magnets. Additional scintillator stations in the spectrometer (shown in yellow) are used to precisely measure the arrival time of a signal with respect to the collision in ATLAS with a precision better than 1 nanosecond. They are also used to trigger the readout of data from the tracking stations.

The final component of the detector is the electromagnetic calorimeter (shown in red). The calorimeter is designed to identify high energy electrons and protons and measure their energies. It is comprised of four spare modules donated by the LHCb experiment, where each module consists of interleaved lead and scintillator plates.

The readout of data from the detector is handled by a trigger and data acquisition system [6]. The data performance and quality depend crucially on this system. As the backgrounds for FASER are expected to be small, the readout of the full detector is triggered by every passage of a detectable high-energy particle. Distinguishing between a signal-like or a background-like event will then be determined in the analysis of the data based on which parts of the detector reported signals.

2 Installation

The components of the FASER detector were assembled and tested in a dedicated area at CERN's Prevessin site. Due to the challenging nature of installing the detector in the tunnel it was vital to perform assembly tests on the surface. This included creating a mock-up of the the base plate that the detector structure is mounted on. The base plate mock-up was assembled on a cement area which included a 1 percent slope to simulate the gradient of the LHC ring. In September 2020 the first assembly was successfully achieved including two magnets, one tracker station, the calorimeter, and all scintillators, as shown in Fig. 3. This was the first time that data from each subdetector was read out simultaneously with the data acquisition system, an exciting milestone in the development of the FASER experiment.

In December 2020, the FASER magnets were then installed in TI12. The rest of the detector was installed in April 2021. Figure 4 shows the TI12 tunnel before and after the installation of the detector. In early November 2021 proton beams were circulated in the LHC for the first time since the LHC Long Shutdown 2 (LS2) started in 2019. During this time there were low intensity proton-proton collisions as well as so-called "beam splash" events, where the proton beams were intentionally directed against targets to create sprays of secondary particles. FASER successfully recorded multiple events containing particles from real proton beams for the first time. The first events with a charged particle passing through all three tracker stations in the spectrometer were also recorded. An example event from real data is shown in Fig. 5. In November 2021 the interface tracker station for the FASERν detector was installed successfully in the tunnel. A photo is shown in Fig. 6. The last component, the emulsion detector FASERν, detector will be installed in March 2022 just before physics collisions will resume in the LHC.

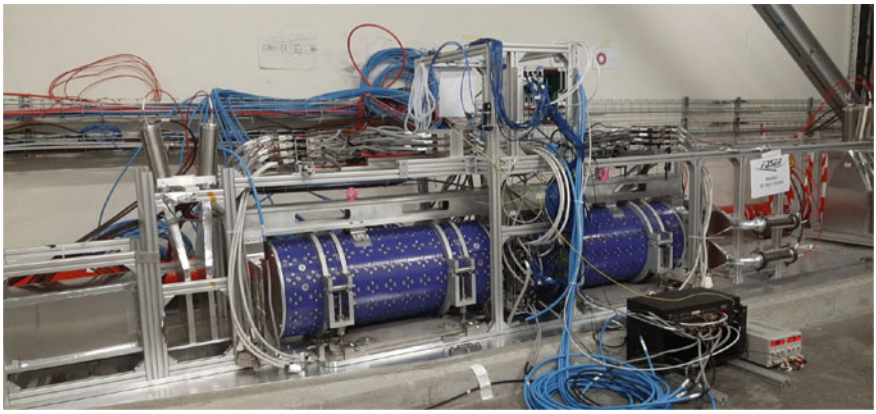

Fig. 3 Photo of part of the FASER detector during the surface assembly test in CERN's Prevessin site in September 2020. The assembly test included two magnets, one tracker station, the calorimeter, and all scintillators

Fig. 4 Top: The TI12 tunnel in December 2019 before civil engineering works and detector installation. Bottom: The FASER detector after installation in the TI12 tunnel in April 2021

3 Outlook

The FASER experiment will be able to explore new lifetime frontiers at the LHC in the search for dark matter. The upcoming LHC run from 2022–2025 will be a very exciting time. In the further future, the proposed Forward Physics Facility (FPF) may provide even further insight into the nature of dark matter at the LHC [7]. It would be located in a similar position to FASER, in the far-forward region of the LHC, in a dedicated experimental cavern. This new cavern would contain diverse experiments such as FLArE, a proposed 10-tonne-scale noble liquid detector which could be used to search for light dark matter particles, and a larger upgraded version of the FASER detector, FASER2.

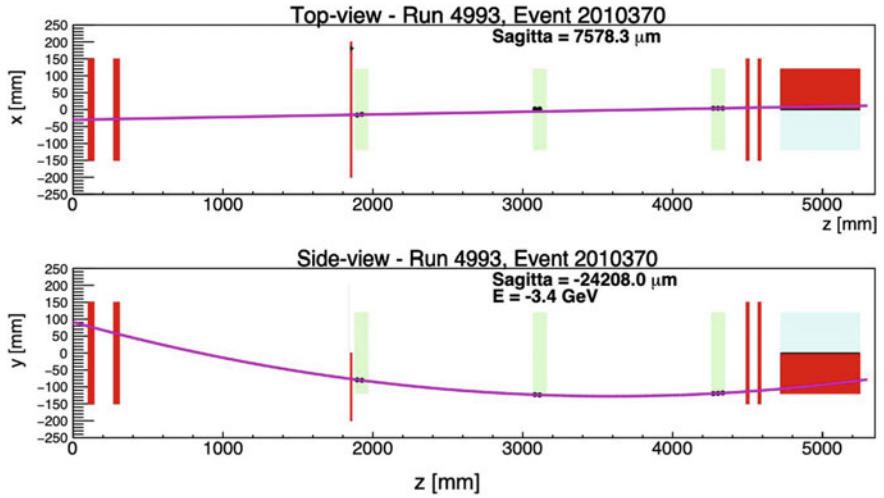

Fig. 5 Event display schematic of a particle track (shown in pink) passing through all three tracker stations (shown in green) in a real event recorded during the LHC pilot beam run in November 2021. Energy deposits in the scintillators (indicated by thick vertical lines) and calorimeters (rectangles on right hand side) are highlighted in red

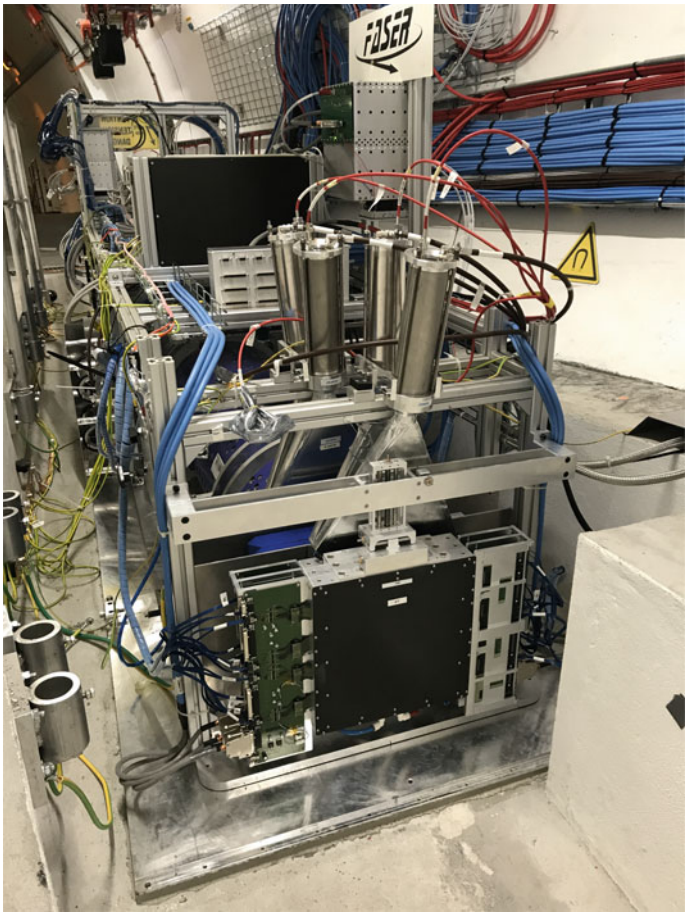

Fig. 6 Photo of the FASERν interface tracker station after installation in the TI12 tunnel in November 2021

References

1. J.L. Feng, I. Galon, F. Kling, S. Trojanowski, ForwArd search experiment at the LHC. Phys. Rev. D **97**(3), 035001 (2018). https://doi.org/10.1103/PhysRevD.97.035001, arXiv:1708.09389 [hep-ph]
2. FASER Collaboration, FASER: ForwArd Search ExpeRiment at the LHC. arXiv:1901.04468 [hep-ex]
3. FASER Collaboration, FASER's physics reach for long-lived particles. Phys. Rev. D **99**(9), 095011 (2019). https://doi.org/10.1103/PhysRevD.99.095011, arXiv:1811.12522 [hep-ph]
4. FASER Collaboration, Detecting and studying high-energy collider neutrinos with FASER at the LHC. Eur. Phys. J. C **80**(1), 61 (2020). https://doi.org/10.1140/epjc/s10052-020-7631-5, arXiv:1908.02310 [hep-ex]
5. FASER Collaboration, The tracking detector of the FASER experiment. arXiv:2112.01116 [physics.ins-det]

6. FASER Collaboration, The trigger and data acquisition system of the FASER experiment. JINST **16**(12), P12028 (2021). https://doi.org/10.1088/1748-0221/16/12/P12028, arXiv:2110.15186 [physics.ins-det]
7. L.A. Anchordoqui et al., The forward physics facility: sites, experiments, and physics potential. arXiv:2109.10905 [hep-ph]

BASE—Testing Fundamental Symmetries by High Precision Comparisons of the Properties of Antiprotons and Protons

Stefan Ulmer

1 Introduction

This chapter reviews the experimental activities of the BASE collaboration at CERN [1], which operates advanced Penning trap systems to compare the fundamental properties of single protons and antiprotons with high precision [2–4]. Such measurements provide stringent tests of charge, parity, time (CPT) reversal invariance, which is the most fundamental symmetry in the standard model of particle physics [5]. CPT invariance has been tested with high sensitivity in very different sectors [6], and so far no indications for a violation have been found. In contrast, the non-observation of primordial antimatter and the matter excess in our Universe are a tremendous challenge for the Standard Model, since the tiny amount of CP-violation contained in the Standard Model is insufficient to reproduce the matter content by more than eight orders of magnitude [7]. Other activities of the BASE collaboration include contributions to searches for axions and axion like particles, light spinless bosons ($m_a \ll 1$ eV$/c^2$) originally proposed to resolve the strong CP problem of quantum chromodynamics. In recent work BASE investigated asymmetric couplings of dark-matter particles to fermions and antifermions [8], and used superconducting resonant single particle detection circuits in strong magnetic fields, to set, in a narrow mass range, competitive limits on the coupling of ALPs to photons [9].

The standard model (SM) of particle physics, the mathematically formalized condensate of human knowledge on the fundamental processes of nature, is the pride and joy of modern physics. Although incredibly successful, the model is known to be incomplete. For example, the SM is to a certain extent phenomenological, since it contains 19 fundamental constants [10]—input parameters which are not determined within the SM itself—and it postulates several mechanisms based on experimental observations, yet unlinked to deeper understanding. Perhaps the greatest triumph of

S. Ulmer (✉)
RIKEN, Ulmer Fundamental Symmetries Laboratory, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
e-mail: stefan.ulmer@cern.ch

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_10

the SM is the prediction of the Higgs-boson, which has been experimentally verified at CERN's LHC in 2012 [11, 12], over 50 years after the prediction of its existence [13, 14]. Other successes include the predictions of the exchange quanta of the weak interaction [15], and of the top and the charm quarks, before the particles were observed. Another great achievement of the SM is the exact calculation of the electron's anomalous magnetic dipole moment represented by its g-factor anomaly of $a = 0.001\,159\,652\,181\,61(23)$ [16], which results in the most precise quantum-electrodynamics (QED) based determination of the fine structure constant: $\alpha = 1/137.035999070(98)$ [17]. Within a slight tension this value is consistent with independent direct measurements [18, 19] that reach a fractional accuracy of 81 parts in a trillion. Frustratingly, the success of the Standard Model is restricted to the description of not more than 5% of the energy content of our universe, and even the fundamental mechanisms to explain how this small fraction came into stable existence have yet to be understood. To date we believe that about 69% of the energy content of the Universe consist of a mysterious substance called dark energy [20], 27% are made out of dark matter [21], and the remaining 5% are made of the baryonic matter forming our planetary system and the celestial bodies in our universe [22]. Not included in the SM is a convincing mechanism that causes the striking imbalance of matter over antimatter that is observed on cosmological scales [7]. Combining the Λ-CDM model and the SM, we would expect to find a radiative universe with a baryon/antibaryon ratio of ≈ 1, and a baryon/photon ratio of 10^{-18}, contradicting the observed ratio of 6×10^{-10} by almost nine orders of magnitude. This observation suggests an asymmetry between matter and antimatter in nature, its origin has yet to be discovered. In principle, matter/antimatter asymmetry can conceptually be explained by means of three requirements first summarized by Sakharov [23], which are C and CP violation, Baryon-number violation, and interactions out of thermal equilibrium. However, the amount of CP violation contained in the Standard Model appears to be insufficient for a convincing explanation of the observed baryon asymmetry [22]. As a consequence, modern physics is currently facing an outstanding challenge, consisting of the tension between excellent quantitative knowledge about the considerable lack of human understanding, contrasted by the SM-counterpart, which is within its limited boundaries incredibly successful.

One strategy to search for physics beyond the SM is to create experimental facts that test, constrain and question its very fundamental assumptions with incredible precision. One of these fundamental corner-stones at the core of each of the flat-space-time relativistic quantum field theories that form the SM, is the discrete charge, parity, time (CPT) reversal invariance. In fact, at the level of known fundamental experimental physics, CPT invariance is the only combination of discrete symmetry transformations that is observed as an exact symmetry of nature [24]. Efforts to place quantum field theory on rigorous mathematical grounds [25] have shown, that CPT symmetry itself requires only very few general assumptions [5], and has therefore a truly fundamental character. Any local (1), Lorentz and translation covariant field theory (2) that is represented by a reasonably smooth field operator implementation (3), and which has a stable vacuum ground state (4) without momentum and angular momentum (5), conserves CPT symmetry. The CPT operation contains charge

conjugation, and defines particular correlations between the fundamental properties of matter/antimatter conjugates. Indeed, as a consequence of CPT invariance, particle/antiparticle-conjugates have identical masses and lifetimes, as well as the same charges and magnetic moments, the later two with opposite sign. Another consequence of CPT invariance is, that conjugate matter and antimatter bound states have exactly the same energy spectrum [5]. Any detected difference in the fundamental properties of matter/ antimatter conjugates would challenge the requirements (1) to (5) and would indicate new physics, which inspires experiments such as BASE, which compare the fundamental properties of matter/antimatter conjugates with great precision.

2 The Standard Model Extension

As a consequence of the generality of CPT symmetry and the non-observation of CPT violating signals in laboratory experiments not many fundamental theories that violate CPT invariance exist. Moreover, similar to CP violation, CPT violation may only occur in very few physical systems. Inspired by the motivation to provide a theoretical framework for establishing quantitative bounds on CPT and Lorentz violating experimental signatures in a model-independent way [26] and in different systems, Kostelecky and collaborators have developed the Standard Model Extension (SME) [27]. This effective field theory, constructed based on the discovery that in string theory mechanisms can occur that could cause spontaneous breaking of Lorentz and CPT symmetry [28], different plausible extensions are introduced into the Standard model Lagrangian, that violate CPT and Lorentz symmetry. Their interaction strengths are characterized by effective phenomenological coupling constants of higher dimensional processes that couple to the fermion fields of the SM. The basic idea of the construction is that any CPT and Lorentz violating effect of higher dimension physics enters the resulting effective field theory as a perturbation

$$\Delta\mathcal{L} = \frac{\lambda}{M^k}\langle T\rangle\bar{\psi}\Gamma(i\partial)^k\psi. \tag{1}$$

Here, λ is an effective coupling constant suppressed by the mass dimension $1/M^k$ of the scale at which the exotic physics occurs, $(i\partial)^k$ represents k four-derivatives acting onto the involved fermion fields, and Γ is some gamma-matrix structure. The term $\langle T\rangle$ represents the non-zero expectation value of a function of Lorentz tensors from some higher dimension CPT breaking mechanisms. By careful consideration of different possible shapes of $\Delta\mathcal{L}$, the effective low-energy field theory remains translationally invariant and covariant under changes of the inertial frame of the observer, but violates CPT and partially breaks covariance under particle boosts. The resulting effective field theory contains spontaneous CPT breaking but features properties like microscopic causality and renormalizability.

The SME couples to different systems with different coefficient strength and contains, due to the scaling with the order of perturbation k in principle an infinite but

$1/M^k$-suppressed amount of possible interactions. In its minimal form, containing only operators of mass dimension 4, the quantum electrodynamics (QED) part of the SME gives an effective modified Dirac equation that reads

$$(i\gamma^\mu D_\mu - m - a_\mu\gamma^\mu - b_\mu\gamma_5\gamma^\mu - \frac{1}{2}H_{\mu\nu}\sigma^{\mu\nu} + ic_{\mu\nu}\gamma^\mu D^\nu + id_{\mu\nu}\gamma_5\gamma^\mu D^\nu)\psi = E\psi, \quad (2)$$

where $iD_\mu = \partial_\mu - qA_\mu$ and A_μ is the electromagnetic potential. The added interactions can be understood as stationary background fields permeating the Universe. The terms $a_\mu\gamma^\mu$ and $b_\mu\gamma_5\gamma^\mu$ violate Lorentz and CPT symmetry. The hypothetical interaction $a_\mu\gamma^\mu$ would shift the reference energies of particle-antiparticle conjugate systems, not affecting the wave-lengths of transition spectra. The term $b_\mu\gamma_5\gamma^\mu$ looks like a pseudo-magnetic field that couples to the spin of particles, in experiments on matter-antimatter conjugates, one of the b_μ-components would lead to a measured stationary splitting of the values of the magnetic moments of the particles. Effects of the related interaction would also lead to different splittings in the spectra of the ground state hyperfine splitting of hydrogen and antihydrogen [29, 30]. The last three terms only violate Lorentz symmetry. All these tensor-like background fields of the SME would also cause annual and sidereal variations in the experimental results, since the Earth rotates and revolves on its orbit around the sun, continuously changing the orientation of the experimental laboratories with respect to the SME-fields. So far, such signals have not been detected [6].

Of particular interest is that all the terms in the SME-modified QED Lagrangian have mass dimension, and the exotic physics mediated by these terms lead to absolute energy shifts of the electromagnetic system $i\gamma^\mu D_\mu - m$ which forms the basis for the respective CPT invariance tests. A consequence of the absolute nature of the corrections is that absolute measurement precision and absolute energy resolution, rather than fractional measurement accuracy, is the appropriate figure of merit to quantify the sensitivity of a CPT invariance test.

3 CERN's Antimatter Program and Other Tests of CPT Invariance

The BASE experiment is located in CERN's AD/ELENA complex [31], sharing this facility with 5 additional collaborations, AEgIS [32], ALPHA [33], ASACUSA [34], GBAR [35], and PUMA [36]. AEgIS, GBAR, and a part of the ALPHA collaboration, are testing the weak equivalence principle by investigating the behaviour of antihydrogen in the gravitational field of the earth. The ASACUSA collaboration is focusing on tests of CPT invariance by ground-state-hyperfine spectroscopy in a beam of polarized antihydrogen atoms produced in a CUSP trap [37]. Using a hydrogen beam in the same apparatus, the collaboration has measured the ground state hyperfine transition frequency $\nu_{\mathrm{GSHFS}} = 1\,420\,405\,748.4(3.4)(1.6)$ Hz, which

has a fractional precision of 2.7 p.p.b., and constitutes the most precise measurement in a beam apparatus [38]. Another initiative within ASACUSA is performing high-resolution spectroscopy of the exotic atom antiprotonic helium [39], in which one of the electrons in the He shell is replaced by an antiproton. Laser spectroscopy on this three body system gives access to the antiproton-to-electron mass ratio $m_{\bar{p}}/m_e$. Using buffer-gas-cooled antiprotonic helium atoms at a temperature of $\approx$1.6(1) K, ASACUSA managed to determine the ratio $m_{\bar{p}}/m_e = 1\,836.152\,673\,4(15)$, which has a fractional precision of 8×10^{-10} and agrees on this level with recent proton-to-electron mass ratio values extracted from laser spectroscopy [40] and Penning trap experiments [41]. The ALPHA collaboration is performing precision measurements on the fundamental properties of antihydrogen ($\bar{H}$) using an atom trap. After a development period of about 10 a, which required the invention of several innovative plasma manipulation techniques, the collaboration reported in 2010 the first successful demonstration of $\bar{H}$-trapping [42]. Based on this experimental success and by steadily developing their apparatus to higher trapping efficiency [43], ALPHA has meanwhile studied the charge neutrality of $\bar{H}$ with a record-precision at the level of 7×10^{-10} [44], and demonstrated in a path-finder experiment a first low-resolution test of the free-fall weak equivalence principle [45]. Most importantly, ALPHA has demonstrated the interaction of $\bar{H}$ with laser light in 2017, and based on this achievement and using 2-photon laser spectroscopy, they have measured the 1S/2S transition in $\bar{H}$ with a fractional resolution of 2 parts per trillion [46]. To date, this measurement provides one of the most precise tests of CPT invariance. The recent demonstration of laser cooling of antihydrogen [47] heralds measurements of this transition frequency with much improved resolution, with the potential to reach precision goals comparable to levels that were achieved with hydrogen [48], where in pioneering experiments measurements with fractional resolutions at the level of $\approx 10^{-15}$ were performed. Note that the frequencies of optical transitions in antihydrogen are mainly defined by the fundamental properties of the positron and its Coulomb interaction with the antiproton. Transition frequencies scale with the reduced mass of the bound antiproton/positron system. For that reason, the sensitivity of optical antihydrogen spectroscopy with respect to CPT-odd physics that would couple to the mass of the antiproton is suppressed by a factor of 1837. On the other hand, antihydrogen experiments are sensitive to interactions, and constrain many more SME coefficients than antiproton-in-trap studies [49].

The experiments at the antimatter facility of CERN stand in a long history of tests of CPT invariance and the SME that have been conducted in very different sectors. A widely recognized CPT-test was carried out by comparing decay channels of neutral K_0 and $\bar{K}_0$ mesons to charged and neutral pions [50]. Thereby, their relative mass difference was measured to be less than 10^{-18}, constraining SME-coefficients to $<10^{-18}$ GeV. Pioneering experiments in Penning traps compared the electron and positron g-factors with fractional precisions on the level of 4 parts per billion [51], reaching SME coefficient bounds $<6 \times 10^{-25}$ GeV, and by comparing the anomaly frequencies of μ^+ and μ^- in a storage ring, another widely recognized test of CPT symmetry was performed [52] that constrains SME coefficients to $<2 \times 10^{-24}$ GeV. Although CPT invariant, the result of this measurement is by 4.2 standard errors

inconsistent with SM predictions. Currently discussed controversially [53], this result inspired the development of a several theories that go beyond the SM, including single field extensions for instance by dark photon, dark Z, two-Higgs doublet, and scalar leptoquark coupling, two-field extensions by e.g. coupling to vector-like leptons, and various supersymmetry (SUSY) scenarios [54].

Also conducted at CERN, but before the AD program at was launched, from the late 1980ies until the LEAR program was stopped, were high precision comparisons of the antiproton-to-proton charge-to-mass ratios. In these experiments by the TRAP collaboration a fractional precision at the level of 90 p.p.t. was achieved [55]. It is worthwhile noting that within these experiments trapping of antiprotons from an accelerator, electron cooling of antiprotons, and resonant non destructive frequency measurements were demonstrated for the first time. All these methods pioneered by the TRAP collaboration are meanwhile part of the basic methods-set applied by CERN's antimatter community and constitute groundbreaking contributions for the development of the current antimatter program.

4 BASE—Penning Traps and Measurement Concepts

BASE [1] is a high precision experiment that relies on frequency measurements on single particles in an advanced cryogenic Penning trap system [56]. A Penning trap consists of a strong, homogeneous magnetic field B_0, in cylindrical coordinates (z, ρ) aligned along the z axis. This field confines particles with mass m and charge q in the radial ρ direction. To prevent the particles from escaping along the magnetic field lines, an electrostatic quadrupole potential $\Phi(z, \rho) = V_0 \left(z^2 - \rho^2/2\right)$ is superimposed. A particle in such crossed electro- and magneto-static fields describes a trajectory, which is composed of three independent harmonic oscillator motions, the modified cyclotron oscillator and the magnetron oscillator at ν_+ and ν_- perpendicular to the magnetic field, and the axial oscillator ν_z along the magnetic field lines. The movement of a trapped charged particle induces femto-amp image currents in the trap electrodes, which are picked up by means of ultra sensitive supercondcting single particle detection systems [57, 58], consisting of superconducting tuned circuits and ultra low noise amplifiers [57]. Recording the detector's output signals while the particles are tuned to resonance, and performing a frequency analysis on the recorded signals allows the determination of the ν_+, ν_-, and ν_z frequencies. Application of the invariance theorem [56]

$$\nu_c^2 = \nu_+^2 + \nu_z^2 + \nu_-^2 \tag{3}$$

relates the trap oscillation frequencies to the free cyclotron frequency $\nu_c = (q\,B_0)/(2\pi m)$ and gives access to the fundamental properties q and m of the single trapped particle. Comparing cyclotron frequencies of protons p and antiprotons $\bar{\text{p}}$, in the same magnetic field, gives access to ratios of charge-to-mass ratios

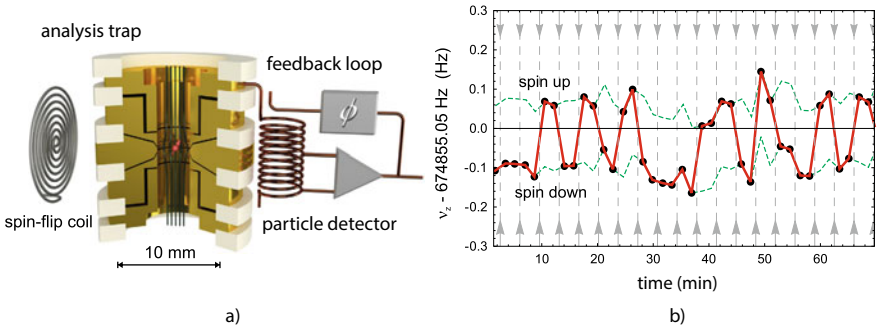

Fig. 1 **a** Schematic illustration of a Penning trap with a superimposed strong magnetic bottle to detect spin transitions by application of the continuous Stern Gerlach effect. **b** first non-destructive observation of single antiproton spin transitions with high fidelity

$$\frac{\nu_{c,\bar{\mathrm{p}}}}{\nu_{c,\mathrm{p}}} = \frac{(q/m)_{\bar{\mathrm{p}}}}{(q/m)_{\mathrm{p}}}, \tag{4}$$

which is the basic concept of any high precision mass spectrometry experiment.

Advanced Penning traps allow, in addition to the determination of cyclotron frequencies ν_c, the measurement of the spin precession or Larmor frequency ν_L of single particles. To this end the continuous Stern-Gerlach effect [59] is applied. Here, a strong quadratic magnetic inhomogeneity $B_z = B_0 + B_2 z^2$ is superimposed on the trap, as illustrated in Fig. 1a. This "magnetic bottle" couples the spin of a particle to its axial oscillation frequency ν_z, which becomes

$$\nu_z = \nu_{z,0} \pm \frac{\mu_p B_2}{2m\nu_z}, \tag{5}$$

where μ_p is the particle's magnetic moment. By measuring the particle's axial frequency while inducing spin transitions by means of magnetic radio-frequency drives injected to the trap at different drive frequencies ν_{rf}, give access to the spin flip probability as a function of the applied drive frequency $P_{\mathrm{SF}}(\nu_{\mathrm{rf}})$, from which the Larmor frequency is obtained [56]. Determining the Larmor frequency ν_{L} together with synchronous measurements of the cyclotron frequency ν_c provides the magnetic g-factor

$$\frac{g_{\mathrm{p}}}{2} = \frac{\nu_L}{\nu_c} = \frac{\mu_{\mathrm{p}}}{\mu_{\mathrm{N}}}, \tag{6}$$

being the magnetic moment μ_{p} of the particle in units of the nuclear magneton μ_{N}. To apply this principle to directly measure the magnetic moment of the proton-antiproton system is a considerable challenge. While this concept has been applied with great success to determine electron/positron magnetic moments [51], as well

as the magnetic moment of the electron bound to highly charged ions [60], the proton/antiproton magnetic moment is about 658 times smaller than that of the electron. In addition, the sensitivity of the detection signal—the axial frequency shift induced by a spin transition—scales inversely proportional to the particle mass which further complicates detection. A sequence of pioneering developments which include the design of the smallest Penning trap ever operated [61] in a precision experiment, with the largest magnetic bottle ever superimposed to a single trapped particle [62], and the development of sensitive single particle detectors with world leading quality [58] eventually led to the first successful observation of single proton spin transitions [62]. This success paved the path towards several world record measurements of the proton magnetic moment [3, 63]. The successful application to the antiproton, Fig. 1b shows the first single antiproton spin transitions ever observed [59], led to thousand-fold improved magnetic moment based tests of CPT invariance, as described below.

5 The BASE Trap Stack

The heart of the BASE experiment is the trap stack, which consists of four cylindrical Penning traps, as shown in Fig. 2. The trap electrodes are made out of oxygen free copper, to prevent oxidation the electrodes are gold-plated. This trap stack is placed in a cryopumped vacuum chamber with a volume of $\approx$1.2 l, in which a vacuum $\approx 10^{-18}$ mbar is achieved. These extreme vacuum conditions allow antimatter storage times on the order of years [64]. The stack consists of the

- **Reservoir Trap:** This trap, proposed in 2011 and successfully implemented in 2014 [65], is the homogeneous most upstream trap. It catches a cloud of typically 100–500 antiprotons, after catching the apparatus is decoupled from the accelerator. Methods were developed to extract single antiprotons from the reservoir [65],

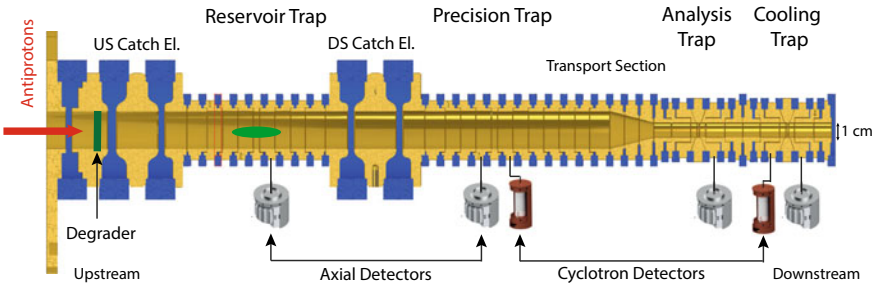

Fig. 2 Schematic of the BASE Penning trap stack. The advanced Penning trap system consists of four trap, a reservoir trap, a precision trap, an analysis trap and a cooling trap. Antiprotons are injected from the upstream side of the instrument. Not shown on the Figure is an electron gun located on the downstream side. All traps are equipped with sensitive single particle detectors

and BASE has demonstrated antiproton storage for more than 405 d. Thanks to this trap, antiproton experiments can be conducted all time of the year, also during accelerator shutdown periods when magnetic background noise in the experiment zone is low. Especially experiment campaigns conducted in these calm periods enable measurements with ultra-high precision.

- **Precision Trap:** The precision trap features an ultra stable, highly homogeneous magnetic field, and frequency tuneable superconducting single particle detectors with world leading quality [58]. All frequency measurements that contribute to high precision results [2, 4] are carried out in this trap.
- **Analysis Trap:** The ring electrode of the analysis trap, with an inner diameter of only 1.8 mm is made out of ferromagentic Co/Fe alloy. This produces in the center of the trap a magnetic bottle with a strength of $\approx$270 kT/m^2. In this trap the continuous Stern Gerlach Effect to determine the spin precession frequency ν_L is applied.
- **Cooling Trap:** The cooling trap is another small trap which provides fast sub-thermal cooling cycles to prepare single particles at cyclotron-mode temperatures $<$100 mK in only a few 10 s, which is essential for magnetic moment measurements at high sampling rate.

All traps are equipped with an ultra sensitive superconducting detector to pick-up the oscillation of the axial mode, the precision trap and the cooling trap have in addition detection systems for direct pickup of the cyclotron mode and its resistive cooling.

6 Antiproton Lifetime

As a consequence of CPT invariance it is required that the lifetimes of the antiproton and the proton, $\tau_{\bar{p}}$ and τ_p, are identical. While the lifetime of the proton has been constrained to $\tau_p > 10^{34}$ a [66], limits on the antiproton lifetime $\tau_{\bar{p}}$ are much lower. They come either from model-dependent comparisons of measured cosmic-ray antiproton

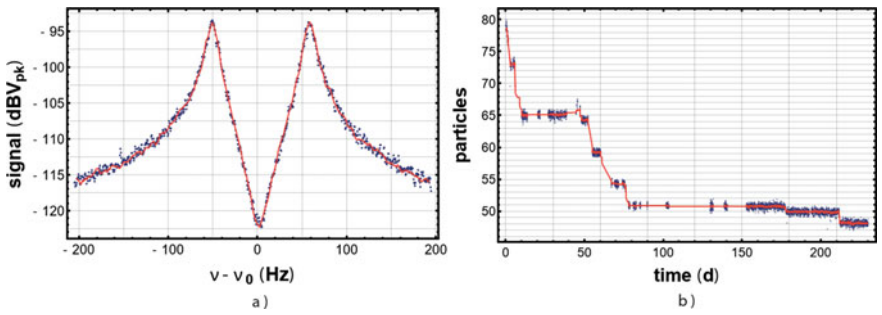

Fig. 3 **a** A cloud of 75 antiprotons tuned to resonance with the axial detection system of the reservoir trap. **b** Content of the reservoir trap as a function of time

flux with cosmic antiproton creation models ($\tau_{\bar{p}} > 800\,000$ a) [66], and from Fermi-Lab's storage ring-based APEX experiment [67], which placed limits on thirteen charged leptonic antiproton decay modes. These experiments constrain the antiproton lifetime, depending on the decay channel, in ranges of 100–700 000 a. Some antiproton decay channels which have been favoured by Grand Unified Theories, such as for example $\bar{p} \rightarrow \nu_{e,\mu} K^-$ [68], have not been studied by the above experiments. The BASE antiproton reservoir-trap provides an elegant and direct way to constrain those, and any yet unexpected antiproton decay channel, based on the direct observation of the cloud of particles which is stored in the trap. To count antiprotons, we use the interaction between the axial detection system of the reservoir trap and the trapped particles. Particles in thermal equilibrium with the detection system which are tuned to resonance with the detector create a short on the fast Fourier transform spectrum of the time transient of the detector's output, as shown in Fig. 3a. The width of this dip signature scales proportional to the number N of trapped particles. Continuous sampling of the trap content as a function of time enables us to derive limits on the antiproton lifetime. An example of the measured particle content of the trap is shown in Fig. 3b. All the observed steps are related to particle extraction and deliberate experimental intervention. Combining all the reservoir runs available since the BASE program was started in 2014, an equivalent single-particle exposure of $T_{\mathrm{exp}} = 33.03$ a is obtained, from which a lower antiproton lifetime limit

$$\tau_{\mathrm{p,lower}} > 26.15\,\mathrm{a} \tag{7}$$

at 68% confidence level is derived. This improves the previous best limit based on direct observation by about a factor of 100. From the constraint we also obtain upper partial pressure limits for hydrogen $p_{\mathrm{H}} < 0.46 \cdot 10^{-18}$ mbar and Helium $p_{\mathrm{He}} < 1.04 \cdot 10^{-18}$ mbar, which is the best ever characterized pressure in a laboratory experiment on earth.

7 Charge-to-Mass Ratio Measurements and Tests of the Weak Equivalence Principle

Proton/antiproton charge-to-mass ratio comparisons provide stringent tests of CPT invariance and tests of the weak equivalence principle [69, 70]. It has been pointed out by Greenberg, that for flat space-time QFTs the detection of dissimilarities in the masses of matter-antimatter conjugates would imply non-local physics beyond the standard model [71]. In addition, detected deviations could arise from different couplings of matter and antimatter conjugates to gravity [72, 73].

To compare proton and antiproton charge-to-mass ratios, the cyclotron frequencies of antiprotons and negatively charged hydrogen ions H^- are measured. These particles serve as perfect proxies for protons, as first demonstrated and applied in [55]. The H^- mass is related to that of the proton by

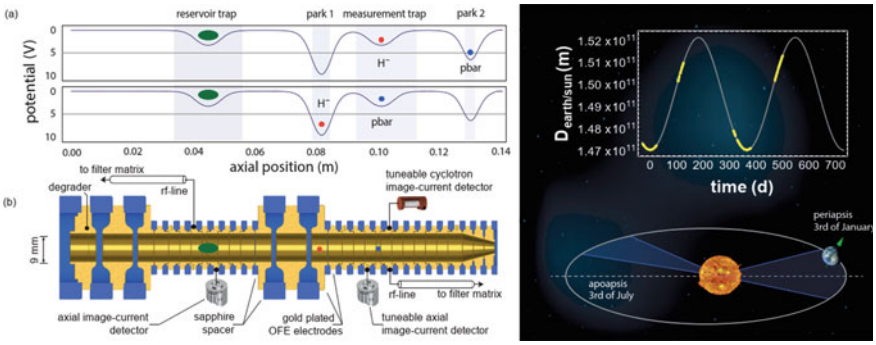

Fig. 4 **a** and **b**: Illistration of the method to measure the proton-to-antuproton charge-to-mass ratio by shuttling antiprotons and H^- ions to the measurement trap. Right figure: Illustration of the test of the clock weak equivalence principle

$$m_{H^-} = m_p \left(1 + 2\frac{m_e}{m_p} - \frac{B_e}{m_p} - \frac{A_e}{m_p} + \alpha_{H^-}\frac{B^2}{m_p}\right). \tag{8}$$

Here m_e/m_p is the electron-to-proton mass ratio, B_e/m_p is the binding energy of the electron in hydrogen, and A_e/m_p is the affinity energy of the second electron in the electron singlet, both in equivalent proton mass units. The term $\alpha_{H^-}B^2/m_p$ is caused by a dynamical frequency shift [74], related to the electrical polarizability α_{H^-} of the H^- ion. All the required corrections are known to a precision level which allows to determine

$$R_{p,H^-,\text{theo}} = 1.001\,089\,218\,753\,80(3)\,, \tag{9}$$

a value with an uncertainty at the level of 0.03 ppt, limited by the current uncertainties in the electron to proton mass ratio [60], the affinity energy of the second bound electron [75] and the polarizability shift [76].

To perform such charge-to-mass ratio comparisons experimentally, BASE has developed a method which allows the accumulation of one frequency ratio measurement in about 240 s, more than 50 times faster than in previous measurements [55]. This technique is based on the simultaneous operation of two traps and two park electrodes, using the concept illustrated in Fig. 4a and b. A cloud of antiprotons and H^- ions is loaded into the reservoir trap [65], from which single particles are extracted to the measurement trap and adjacent park electrodes, shown on the right of Fig. 4. By applying potential ramps to the electrodes, the particles are alternatingly shuttled to the center of the measurement trap, where the cyclotron frequencies are determined. By performing such experiments over a time window of ≈1.5 a, about 24000 frequency ratio measurements were performed, giving the antiproton-to-proton charge-to-mass ratio

$$R_{\bar{p},p,\text{exp}} = -1.000\,000\,000\,003\,(16)\,. \tag{10}$$

The result has an experimental uncertainty of 16 p.p.t. (C.L. 0.68), supporting CPT invariance. It improves the previous measurement [2] by a factor of 4.3 and upon earlier results [55] by a factor of 5.6.

In an illustrative model, that can however not be trivially incorporated into relativistic quantum field theory [77], Hughes and Holzscheiter have shown [70] that if there was a scalar- or tensor-like gravitational coupling to the energy of antimatter that violates the $\mathrm{WEP_{cc}}$ [69], there will be, at the same height in a gravitational field, a frequency difference

$$\frac{\nu_{c,\bar{\mathrm{p}}} - \nu_{c,\mathrm{p}}}{\nu_{c,\mathrm{avg}}} = \frac{3\Phi}{c^2}\left(\alpha_g - 1\right) \tag{11}$$

between a proton cyclotron-clock at $\nu_{c,\mathrm{p}}$ and its CPT conjugate antiproton clock at $\nu_{c,\bar{\mathrm{p}}}$. Here $\alpha_g - 1$ is a parameter characterizing the strength of the potential $\mathrm{WEP_{cc}}$ violation and Φ the gravitational potential. Together with the gravitational potential of the local supergalactic cluster ($\Phi/c^2 = (\mathrm{G\,M})/(rc^2) = 2.99 \cdot 10^{-5}$) [78, 79], the measurement constrains those $\mathrm{WEP_{cc}}$ violating gravitational anomalies to a level of $|\alpha_g - 1| < 1.8 \cdot 10^{-7}$, improving the previous best limits by about a factor of 4. This approach has been discussed controversially [72], since the imposed clock shift depends on the absolute value of the gravitational potential, and a WEP-violating force might have a finite range which would modify the chosen potential. This inspires the following analysis: We use the amplitude $d\Phi_{\mathrm{S}}$ of the change of the mean gravitational potential $\Phi_{\mathrm{S,avg}}$ at the location of our experiment, which is sourced by the earth's elliptic orbit $O(t) = D_p \cdot (1-\epsilon^2)/(1+\epsilon\cos((2\pi/t_{\mathrm{sid}})t))$—with eccentricity $\epsilon = 0.017$, perihelion distance D_p, and time of the sidereal year t_{sid}—around the sun. The eccentricity ϵ leads to a fractional peak-to-peak variation of $2d\Phi_{\mathrm{S}}/\Phi_{\mathrm{S,avg}} \approx 0.03$, as shown in Fig. 4, right. In case of WEP violation, this would induce a cyclotron frequency ratio variation

$$\frac{\Delta R(t)}{R_{\mathrm{avg}}} = \frac{3GM_{\mathrm{sun}}}{c^2}\left(\alpha_{g,D} - 1\right)\left(\frac{1}{O(t)} - \frac{1}{O(t_0)}\right), \tag{12}$$

where M_{sun} is the mass of the sun and G the gravitational constant. As shown in Fig. 4, the acquired data is distributed such that about 80% of the total available peak-to-peak variation of Φ_S are covered. Projecting the measured frequency ratios to one sidereal year, and looking for oscillations of the measured frequency ratio constrains $|\alpha_{g,D} - 1| < 0.030$ (C.L. 0.68). This limit is similar to the initial goals of model-independent experiments testing the weak equivalence principle $\mathrm{WEP_{ff}}$ by dropping antihydrogen in the gravitational field of the earth [35, 80, 81].

At the currently quoted uncertainty of 16 p.p.t., this measurement also provides a 4-fold improved limit on the coefficient r^{H^-} of the minimal SME [6, 82], becoming $r^{H^-} < 2.09 \cdot 10^{-27}$ (C.L. 0.68). Very recently an additional non-minimal extension of the SME with up to mass dimension six was applied to experiments that particle/antiparticle charge-to-mass ratios [83] in Penning traps. The related charge-to-mass ratio figure of merit

Table 1 Constraints on coefficients of the standard model extension. The second column describes the previous best limit based on [55] and [2], theorized and summarized in [83]. The third column gives the improved limit based on the measurement presented here, the fourth column shows the ratio of the second and the third column. All entries are based on C.L. 0.68

Coefficient	Previous Limit	Improved Limit	Factor
$\lvert\tilde{c}_e^{XX}\rvert$	$<3.23 \cdot 10^{-14}$	$<7.79 \cdot 10^{-15}$	4.14
$\lvert\tilde{c}_e^{YY}\rvert$	$<3.23 \cdot 10^{-14}$	$<7.79 \cdot 10^{-15}$	4.14
$\lvert\tilde{c}_e^{ZZ}\rvert$	$<2.14 \cdot 10^{-14}$	$<4.96 \cdot 10^{-15}$	4.31
$\lvert\tilde{c}_p^{XX}\rvert, \lvert\tilde{c}_p^{*XX}\rvert$	$<1.19 \cdot 10^{-10}$	$<2.86 \cdot 10^{-11}$	4.14
$\lvert\tilde{c}_p^{YY}\rvert, \lvert\tilde{c}_p^{*YY}\rvert$	$<1.19 \cdot 10^{-10}$	$<2.86 \cdot 10^{-11}$	4.14
$\lvert\tilde{c}_p^{ZZ}\rvert, \lvert\tilde{c}_p^{*ZZ}\rvert$	$<7.85 \cdot 10^{-11}$	$<1.82 \cdot 10^{-11}$	4.31

$$|\delta\omega_c^{\bar{p}} - R_{\bar{\mathrm{p}},\mathrm{p},\mathrm{exp}}\delta\omega_c^{p} - 2R_{\bar{\mathrm{p}},\mathrm{p},\mathrm{exp}}\delta\omega_c^{e^-}| < 1.96 \times 10^{-27}\,\mathrm{GeV}, \tag{13}$$

where $\frac{\delta\omega_c^w}{q_0 B}$ is a function of coefficients $\tilde{b}_w$ and $\tilde{c}_w$ that describe the strengths of feebly interacting CPT-violating background fields, coupling to particles w, the antiproton $\bar{\mathrm{p}}$, the proton p, and the electron e^-, enables us to set improved limits on the nine SME coefficients that are summarized in Table 1.

8 Magnetic Moment Measurements

Proton and antiproton magnetic moment measurements are among the most difficult experiments that can be carried out in Penning traps, since they require the simultaneous measurement of the spin precession frequency ν_L and the cyclotron frequency ν_c to determine the magnetic moment $g/2 = \mu_p/\mu_N = \nu_L/\nu_c$. While the measurement of the cyclotron frequency is a straight-forward technique, the determination of ν_L is a challenge. As mentioned above, the sensitivity of the continuous Stern Gerlach effect which is used to get access to ν_L scales proportional to the magnetic moment over mass ratio. Compared to the electron, this ratio is for protons more than 1 million times smaller, which makes proton-antiproton spin transitions extremely difficult to observe. Given technical constraints of state of the art Penning trap experiments, the detection of antiproton spin transitions requires the use of a magnetic bottle with a strength of the order of $B_2 = 100\,\mathrm{kT/m^2}$, in BASE $B_2 = 276(8)\,\mathrm{kT/m^2}$ is used. At the parameters of the experiment a spin transition induces in such a B_2 an axial frequency shift of about 170 mHz out of ≈650 kHz. To observe these tiny frequency shifts induced by spin transitions under the extreme magnetic conditions is challenging, since noise-driven quantum fluctuations in the radial modes induce considerable axial frequency instabilities [84]. Thankfully, the cyclotron transition rate scales with the radial quantum number of the cyclotron oscillator, such that for radially cold particles, with cyclotron energies of the order 10 μeV, high fidelity spin

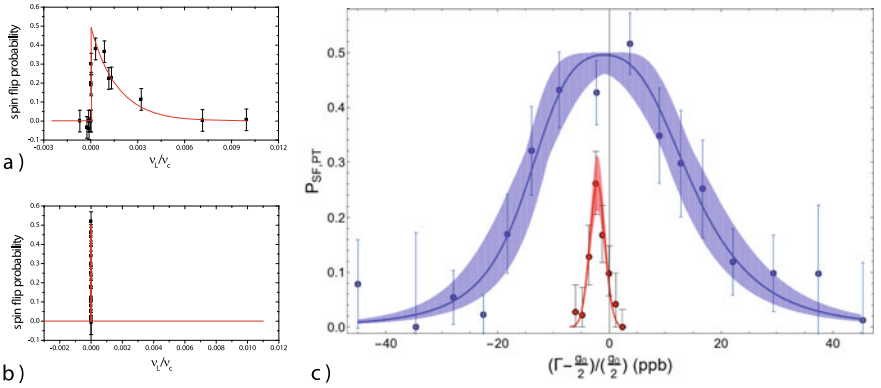

Fig. 5 Antiproton g-factor resonances measured by BASE. **a** 2016 g-factor resonance measured in the BASE analysis trap with its strong superimposed magnetic bottle. **b** g-factor resonance measured with the two-particle/three-trap method plotted on the same scale as (**a**). **c** zoomed version of the antiproton (blue) g-factor resonance shown in (**b**) and proton (red) g-factor resonance measured in 2016/2017 at BASE-Mainz

transition detection can be achieved [59, 62]. To reach such low temperatures, a method called sub-thermal cooling is used, which provides low radial particle temperatures, the method is however time consuming to apply. Another disadvantage of the strong magnetic bottle is that it leads to broadening of the Larmor resonance line, which limits single trap magnetic moment measurements to the parts per million level [85]. To overcome these limitations and inspired by ideas of Haeffner and collaborators [86], we invented a two particle/three trap technique described in detail in [4], which measures the magnetic moment of the particle in the homogeneous magnetic field of the precision trap, here the magnetic bottle trap is solely used as a spin state analyzer. Compared to measurements in the inhomogeneous analysis trap this narrows down the spin line by more than a factor of 1000, as shown in Fig. 5 a and b. The successful implementation of this technique enabled the determination of the antiproton magnetic moment to

$$\frac{\mu_{\bar{p}}}{\mu_N} = -2.792\,847\,3443(46), \tag{14}$$

which has a fractional resolution of 1.7 ppb, and improves the previous best measurement carried out by the ATRAP collaboration [87] by more than a factor of 3000. Measurements carried-out in parallel at the BASE Mainz experiment [3] determine the proton magnetic moment to

$$\frac{\mu_p}{\mu_N} = 2.792\,847\,344\,62(82). \tag{15}$$

This improves the previous best measurement [63] by a factor of 11, and the precision of results based on hydrogen masers in magnetic fields [88] by a factor of 33.

The combination of these two antiproton and proton magnetic moment measurements can be combined to

$$\left(\frac{\mu_{\mathrm{p}}}{\mu_{\mathrm{N}}}+\frac{\mu_{\bar{\mathrm{p}}}}{\mu_{\mathrm{N}}}\right)=0.3(8.3)\times 10^{-9}, \tag{16}$$

the uncertainty in brackets represents the 95% confidence interval. Within the uncertainty of the measurements our results test the standard model at an energy level $<8.1\times 10^{-25}$ GeV and support CPT invariance. These measurement enable us to set constraints on six different exotic CPT and Lorentz violating DC coefficients of the non-minimal Standard Model Extension, with $|\tilde{b}_{\mathrm{p}}^{Z}|<8.1\times 10^{-25}$ GeV, $|\tilde{b}_{\mathrm{F,p}}^{\mathrm{XX}}+\tilde{b}_{\mathrm{F,p}}^{\mathrm{YY}}|<4.6\times 10^{-9}$ GeV, and $|\tilde{b}_{\mathrm{F,p}}^{\mathrm{ZZ}}|<3.3\times 10^{-9}$ GeV for protons, ans $|\tilde{b}_{\mathrm{p}}^{*\mathrm{Z}}|<1.5\times 10^{-24}$ GeV, $|\tilde{b}_{\mathrm{F,p}}^{*\mathrm{XX}}+\tilde{b}_{\mathrm{F,p}}^{*\mathrm{YY}}|<3.1\times 10^{-9}$ GeV, and $|\tilde{b}_{\mathrm{F,p}}^{*\mathrm{ZZ}}|<1.1\times 10^{-9}$ GeV for protons. In addition, our measurements enable us to set limits on a possible magnetic moment splitting mediated by interactions of the form $f_{\mathrm{p}}^{0}B\sigma$ which arise from ultra short distance scale physics [89]. Here we derive $f_{\mathrm{p}}^{0}=\mu_{\mathrm{N}}(g_{\mathrm{p}}-g_{\bar{\mathrm{p}}})/4<4.5\times 10^{-12}$, which improve the previous best constraints by about three orders of magnitude.

9 Constraints on Antimatter/Dark Matter Interaction

The results of the magnetic moment measurements described above can be used to search for asymmetric dark matter/antimatter couplings [8]. If axions or axion-like particles exist, they have been hypothetically produced in the early Universe by non-thermal mechanisms, such as for example vacuum misalignment, and form a coherently oscillating classical field: $a=a_0\cos(\omega_a t)$, where the oscillation frequency is given by the Compton frequency $\nu_a\approx m_a c^2/h$. Here, m_a is the axion mass, c the speed of light and h the reduced Planck constant. The axions may interact with protons and antiprotons by so-called derivative interactions causing spin precession [90]. These interactions would cause sidebands and line shape broadening effects in the magnetic moment resonances sampled by BASE.

In the non-relativistic limit, the relevant part of these interactions can be described by the time-dependent Hamiltonian [91]

$$H_{\mathrm{int}}(t)\approx\frac{C_{\bar{p}}a_0}{2f_a}\sin(\omega_a t)\sigma_{\bar{p}}\cdot p_a\,, \tag{17}$$

where $\sigma_{\bar{p}}$, p_a and $C_{\bar{p}}/f_a$ are the Pauli spin-matrix vector of the antiproton, the axion-field momentum vector, and the axion-antiproton interaction parameter, respectively. The leading-order shift of the antiproton spin-precession frequency due to the above interaction is given by:

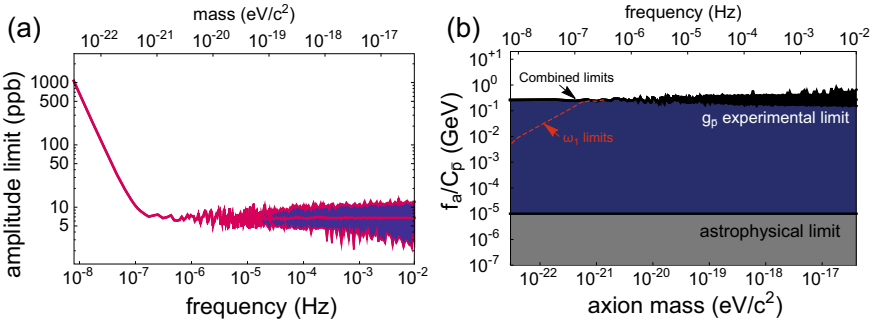

Fig. 6 **a** Upper 95% confidence limits on the oscillation amplitude $b_{\mathrm{up}}(\omega)$ of the antiproton Larmor frequency. **b** 95% confidence limits on the axion-antiproton interaction parameter as a function of the axion mass. The grey area shows the parameter space excluded by axion emission from antiprotons in SN 1987A. The dark blue area shows the parameter space excluded from our analysis of the antiproton spin-flip data. The black line shows the upper limit of the excluded area by using the most significant limit from the main and two sideband modes. The yellow dashed line shows the limits from only detecting at the main frequency ω_1 for $m_a < 10^{-21}\mathrm{eV}/c^2$

$$\delta\omega_L^{\bar{p}}(t) \approx \frac{C_{\bar{p}} m_a a_0 |v_a|}{f_a} \left[A\cos(\Omega_{\mathrm{sid}} t + \alpha) + B\right] \sin(\omega_a t)\,, \tag{18}$$

where $|v_a| \sim 10^{-3}c$ is the average speed of the galactic axions with respect to the Solar System, $\Omega_{\mathrm{sid}} \approx 7.29 \times 10^{-5}\ \mathrm{s}^{-1}$ is the sidereal angular frequency, and $\alpha \approx -25°$, $A \approx 0.63$, and $B \approx -0.26$ are parameters determined by the orientation of the experiment relative to the galactic axion dark matter flux. It is noted note that the time-dependent perturbation of the antiproton spin-precession frequency in Eq. (18) has three underlying angular frequencies: $\nu_1 = \nu_a$, $\nu_2 = \nu_a + \Omega_{\mathrm{sid}}/(2\pi)$, and $\nu_3 = |\nu_a - \Omega_{\mathrm{sid}}/(2\pi)|$, with the power distributed approximately evenly between the three modes.

By analyzing the time sequence data recorded during the antiproton magnetic moment measurement campaign as described in [8], we obtain limits on the asymmetric axion-antiproton coupling coefficient $f_a/C_{\overline{p}} > 0.3\,\mathrm{GeV}$ over the entire mass range $10^{-22}\,\mathrm{eV}/c^2 < m_a < 10^{-16}\,\mathrm{eV}/c^2$ as shown in Figs. 6 and 7. In the considered mass range our laboratory limits are by five orders of magnitude stronger than bounds derived from astrophysical studies.

This slow oscillation analysis allows as well to constrain time dependent coefficients of the CPT and Lorentz violating terms of the standard model extension [92]. Using the limits derived from the time series analysis and considering the orientation of our experiment [85], enables BASE to constrain 6 combinations of time-dependent coefficients of the SME, $|\tilde{b}_p^{*X}| < 2.5 \times 10^{-24}\,\mathrm{GeV}$, $|\tilde{b}_p^{*Y}| < 2.5 \times 10^{-24}\,\mathrm{GeV}$, $|\tilde{b}_p^{*XX} - \tilde{b}_p^{*YY}| < 1.6 \times 10^{-8}\,\mathrm{GeV}^{-1}$, $|\tilde{b}_p^{*XZ}| < 1.0 \times 10^{-8}\,\mathrm{GeV}^{-1}$, $|\tilde{b}_p^{*YZ}| < 1.0 \times 10^{-8}\,\mathrm{GeV}^{-1}$, and $|\tilde{b}_p^{*XY}| < 8.2 \times 10^{-9}\,\mathrm{GeV}^{-1}$.

The work summarized here and originally published in [8] provides the first limits on the axion-antiproton coupling in the 10^{-22} eV to 10^{-16} eV mass range. Similar

constraints could be set for other antiparticles, namely positrons [51] and anti-muons [52], from a frequency-domain analysis of their (g-2) measurements.

10 BASE CDM—Superconducting Haloscope

Another recently developed experimental branch is dedicated to the search for axions and axion like particles (ALPs), operating the experiment as a haloscope based on the ultra sensitive single particle detection systems that are commonly used in BASE. If ALPs exist, they couple to electric fields **E** and magnetic fields **B** fields through the Lagrange density term

$$\mathcal{L}_{a\gamma\gamma} = -g_{a\gamma} a(\mathbf{x}) \mathbf{E}(\mathbf{x}) \cdot \mathbf{B}(\mathbf{x}) \,, \tag{19}$$

where $a(\mathbf{x})$ is the local axion field and $g_{a\gamma}$ is a coupling constant [93, 94]. In a strong magnetic field, such as present in the superconducting magnet of the BASE experiment, this term allows axions to convert into photons. The signals sourced by axions can be picked up by sensitive resonant LC circuits in strong magnetic fields [95]. In [9] we have shown how the superconducting detectors, usually used to pick-up single particle signals [58, 96], can be used as sensitive antennas for ALPs. The superconducting resonant LC circuits are also sensitive to any changes in the magnetic flux within the toroidal inductor, achieving maximum sensitivity when the frequency of those flux changes matches the resonant frequency of the LC circuit. If axion-like particles oscillate through the strong magnetic field of the BASE magnet, they source oscillating magnetic fields which produce exactly these flux changes [95]. To calculate the strength of these fields, we note that the presence of an axion field a modifies Maxwell's equations. In particular in the presence of a strong external magnetic field $\mathbf{B}_e$ and no free electric current density, the Maxwell-Ampere equation becomes

$$\nabla \times \mathbf{B} - \mu_0 \dot{\mathbf{D}} = -g_{a\gamma} \mathbf{B}_e \dot{a} \,, \tag{20}$$

where **B** and **D** are the usual classical fields. By solving Eq. 20 for the boundary conditions of the BASE toroidal detection systems, noting that its dimensions of the superconducting circuit are much smaller than the axion wavelength $\lambda_a = h/(m_a c)$, we find that the right hand side term sources an azimuthal magnetic field $\mathbf{B}_a$, shown in red in Fig. 7a, which oscillates at a frequency $\nu_a = m_a c^2 (1 + v \cdot v/(2c^2))/h$, where v is the axion velocity and

$$\mathbf{B}_a = -\frac{1}{2} g_{a\gamma} r \sqrt{\rho_a \hbar c} |\mathbf{B}_e| \hat{\phi} \,. \tag{21}$$

Here $\rho_a \hbar c = 4\pi^2 \nu_a^2 |a|^2/2$ is the local axion energy density and r is the radial coordinate. The oscillating magnetic field leads to a changing flux in the inductor, which

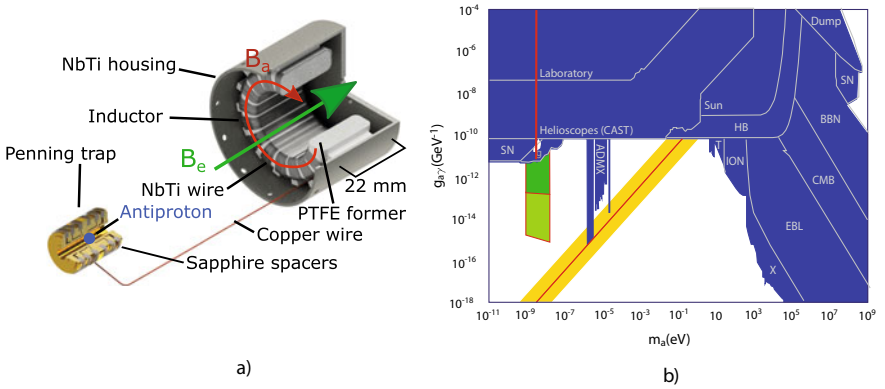

Fig. 7 **a** Superconducting toroidal resonator connected to a Penning trap. The green arrow indicates the external magnetic field of the superconducting magnet, the red arrow indicates the magnetic field sourced by hypothetical axions. **b** limits plot which sets the BASE measurement into context with other quests. With purpose built experiments we expect to constrain the green region

in turn produces an oscillating voltage at the input of the first cryogenic amplification stage connected to the detector. By calibrating the temperature of the detection system using a single particle in the BASE analysis trap, and searching the detector's resonance spectra for narrow peak signatures induced by hypothetical ALPs, we were able to constrain the axion to photon coupling constant $g_{a,\gamma} < 1 \times 10^{-11}$/GeV in the axion mass range around $2.7906\,\text{neV} < m_a c^2 < 2.7914\,\text{neV}$. These constraints are more than one order of magnitude lower than the best laboratory haloscope and approximately 5 times lower than the CERN axion solar telescope (CAST).

To adapt these detectors into more powerful axion search experiments with higher detection bandwidth, we were developing superconducting tunable detectors with which we have demonstrated a more than 1000-fold improved scan bandwidth at improved quality factor. The detection sensitivity of the axion LC experiment scales with the volume of the detection toroid, thus we are currently designing the dedicated experiment BASE-CDM, which covers the entire bore of the superconducting experiment magnet with sensitive axion antennas. With this new instrument, which will also incorporate magnetic field tuning, improved low noise amplifiers and additional mechanisms for frequency tuning, we anticipate in future runs a $\approx$100-fold improved detection sensitivity at $>$1000-fold increased detection bandwidth.

11 Future Technologies

11.1 Transportable Antimatter Traps

The precision measurements of BASE reached for determinations of magnetic moments fractional resolutions on the parts per billion level, and for charge-to-mass ratios measurements resolutions on the level of 16 parts per trillion. In particular the parts per trillion measurements are limited by magnetic field noise in the accelerator hall, it is a challenge to further improve the resolution of these studies in the current environment. Thus we are planning to develop a transportable antiproton trap, which will allow us to relocate antiprotons trapped in the AD/ELENA facility to calm precision laboratory space. We expect that this transportable trap, BASE-STEP, will become operational in 2022–2023 to provide a fruitful perspective for proton/antiproton precision comparisons with steadily improved future precision.

11.2 Laser-Cooling

To achieve high spin-state-detection fidelity in magnetic moment measurements, particles at low cyclotron mode temperature are needed [59]. To date, selective resistive cooling by means of 4 K thermal resistors is used to prepare particles at radial energies $E_+ < 10\,\mu\text{eV}$, which is a time consuming statistical process [85]. To deterministically prepare particles at such low energies, a new experiment was developed, which allows to couple protons to a reservoir of laser cooled $^9\text{Be}^+$ ions. The $^9\text{Be}^+$ ions are cooled to temperatures $<20\,\text{mK}$, the protons are coupled to these particles either using a common endcap technique [97] or by image current interaction mediated by a superconducting tuned circuit [98]. Very recently, dramatic progress was achieved in this experiment, and first sympathetic cooling of protons was demonstrated. In these path-finder experiments the axial temperature of protons was reduced from initially 17 K to a final temperature of 2.6(2.5) K [98]. Theoretical considerations imply that with an optimized experiment and advance coupling procedures, particle temperatures of $\approx 10\,\text{mK}$ can be achieved in proton/$^9\text{Be}^+$ interaction times of 10 s, implying a bright future for proton/antiproton magnetic moment comparisons at improved sampling rate [99].

11.3 Quantum Logic Spectroscopy

As another initiative in BASE, the collaboration is working on an experimental method inspired by quantum logic techniques that could improve particle localization and readout speed. The method will allow sympathetic cooling of protons and antiprotons to the the quantum-mechanical ground state, as well as the quantum-logic

inspired readout of the particle's spin state [100]. Both of these features are achieved through coupling to a laser-cooled 'logic' ion which is co-trapped in a double-well potential. In some aspects similar to the experiment described above, the double well potential will provide much higher coupling rates and will enable more sensitive particle manipulation schemes. This will add an additional boost of experiment sampling rate and reduce systematic uncertainties of future measurement campaigns

12 Summary and Conclusions

Within this chapter the achievements made by the BASE collaboration, which operates advanced Penning trap systems to compare the fundamental properties of protons and antiprotons, were summarized. These experiments test matter/antimatter symmetry and the fundamental CPT invariance at lowest energies and with greatest precision, and test the SM at energy resolutions partly better than 2×10^{-27} GeV. Since the start of the BASE program in 2013, the collaboration improved previous antiproton/proton mass comparisons [55] to a level of 16 parts per trillion in fractional and, using innovative new trap-technologies [4], proton/antiproton magnetic moments were compared with parts per billion precision [3, 4], improving previous constraints [87] by more than three orders of magnitude. These experiments set limits on exotic physics at resolutions comparable to lepton experiments [51, 52], as graphically summarized in Fig. 8. In addition to tests of CPT invariance, BASE contributes to the quest for axions and axion like particles, searching for asymmetric antimatter/dark matter couplings in antiproton magnetic moment resonances [8]. The results of these studies provide orders of magnitude more stringent constraints on asymmetric antimatter/dark matter interactions than astrophysical searches. Very recently BASE has used superconducting single particle detection systems to operate the experiment as an LC-haloscope [9], and has set in a narrow mass range laboratory

Year	Matter $g/2$	Antimatter $\overline{g}/2$	CPT $\lvert g/\overline{g}\rvert - 1$	System	SME $\lvert b_L\rvert$ (GeV)	$\lvert f_X^0\rvert\ (\mu_B)$
1987	1.001 159 652 188 9 (43)	1.001 159 652 187 9 (43)	0.000 000 000 000 5 (21)	electron/positron	$6 * 10^{-25}$	$2 * 10^{-12}$
2006	1.001 165 921 5 (11)	1.001 165 920 4 (12)	0.000 000 001 1 (12)	muon (μ^-, μ^+)	$1 * 10^{-23}$	$3 * 10^{-11}$
2017	2.792 847 344 62 (82)	2.792 847 344 1 (42)	0.000 000 000 2 (15)	proton/antiproton	$8 * 10^{-25}$	$2 * 10^{-12}$

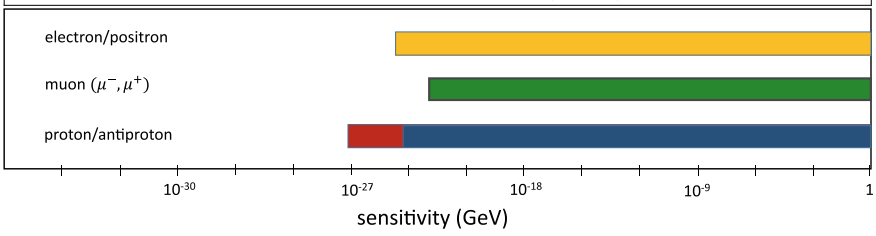

Fig. 8 Measurement precision of CPT tests performed with lepton and baryon magnetic moments and their sensitivity with respect to CPT-odd coefficients of the Standard Model Extension

limits that are about 5 times more stringent than any other laboratory limits in the respective mass range. The limits have similar resolution as satellite based experiments and cover a previously unconstrained region of the current exclusion landscape [9]. The development of a purpose-built experiment, BASE-CDM, to push these limits to higher sensitivity at considerably improved mass bandwidth is under way. In order to steadily improve CPT-invariance-tests in the bayron sector towards improved precision, BASE is currently developing several future technologies, among them new types of Penning trap experiments that use sympathetic cooling of antiprotons and protons, by coupling the particles to laser-cooled $^{9}Be^{+}$ ions [98, 99], and by implementing quantum logic inspired spin state readout [100]. These new trapping technologies will enable CPT tests in the baryon sector at improved sampling rate and precision. It is anticipated that these experiments will reach precision limits defined by the environmental conditions of the accelerator hall. Inspired by these anticipated limits the transportable trap BASE-STEP is currently under development, which will allow BASE to relocate antiprotons from noisy accelerator environment to dedicated calm laboratory space, providing a bright future for proton-antiproton based tests of CPT invariance.

Acknowledgements I acknowledge technical support by CERN, especially the Antiproton Decelerator operation group, CERN's cryolab team and engineering department, and all other CERN groups which provide support to Antiproton Decelerator experiments. In addition I acknowledge financial support by RIKEN, the RIKEN EEE pioneering project funding, the Max-Planck Society, and the Max-Planck, RIKEN, PTB-Center for Time, Constants, and Fundamental Symmetries (C-TCFS). I would like to express my deepest gratitude towards the entire BASE team for all the dedicated work and creativity put into the project. In particular I would like to thank Klaus Blaum, collaborator at BASE and director of the Max Planck Institute for Nuclear Physis, Heidelberg, and Yasunorik Yamazaki, Chief Scientist at RIKEN, for their invaluable support, friendship and encouragement. I thank Christian Ospelkaus for support, friendship, trustful collaboration and his future visions, and in particular Christian Smorra for his dedicated contributions to BASE.

References

1. C. Smorra, K. Blaum, L. Bojtar, M. Borchert, K. Franke, T. Higuchi, N. Leefer, H. Nagahama, Y. Matsuda, A. Mooser, M. Niemann, C. Ospelkaus, W. Quint, G. Schneider, S. Sellner, T. Tanaka, S.V. Gorp, J. Walz, Y. Yamazaki, S. Ulmer, Europ. Phys. J. Special Topics **224**(16), 3055 (2015)
2. S. Ulmer, C. Smorra, A. Mooser, K. Franke, H. Nagahama, G. Schneider, T. Higuchi, S. Van Gorp, K. Blaum, Y. Matsuda et al., Nature **524**(7564), 196 (2015)
3. G. Schneider, A. Mooser, M. Bohman, N. Schön, J. Harrington, T. Higuchi, H. Nagahama, S. Sellner, C. Smorra, K. Blaum et al., Science **358**(6366), 1081 (2017)
4. C. Smorra, S. Sellner, M.J. Borchert, J.A. Harrington, T. Higuchi, H. Nagahama, T. Tanaka, A. Mooser, G. Schneider, M. Bohman et al., Nature **550**(7676), 371 (2017)
5. R. Lehnert, Symmetry **8**(11), 114 (2016)
6. V.A. Kosteleckỳ, N. Russell, Rev. Modern Phys. **83**(1), 11 (2011)
7. M. Dine, A. Kusenko, Rev. Modern Phys. **76**(1), 1 (2003)
8. C. Smorra, Y.V. Stadnik, P.E. Blessing, M. Bohman, M.J. Borchert, J.A. Devlin, S. Erlewein, J.A. Harrington, T. Higuchi, A. Mooser, G. Schneider, M. Wiesinger, E. Wursten, K. Blaum,

Y. Matsuda, C. Ospelkaus, W. Quint, J. Walz, Y. Yamazaki, D. Budker, S. Ulmer, Nature **575**(7782), 310 (2019). https://doi.org/10.1038/s41586-019-1727-9
9. J.A. Devlin, M.J. Borchert, S. Erlewein, M. Fleck, J.A. Harrington, B. Latacz, J. Warncke, E. Wursten, M.A. Bohman, A.H. Mooser et al., Phys. Rev. Lett. **126**(4), 041301 (2021)
10. M. Safronova, D. Budker, D. DeMille, D.F.J. Kimball, A. Derevianko, C.W. Clark, Rev. Modern Phys. **90**(2), 025008 (2018)
11. G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S.A. Khalek, A.A. Abdelalim, R. Aben, B. Abi, M. Abolins, O. AbouZeid et al., Phys. Lett. B **716**(1), 1 (2012)
12. S. Chatrchyan, V. Khachatryan, A.M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, M. Friedl et al., Phy. Lett. B **710**(1), 26 (2012)
13. P.W. Higgs, Phys. Rev. Lett. **13**(16), 508 (1964)
14. F. Englert, R. Brout, Phys. Rev. Lett. **13**(9), 321 (1964)
15. S. Weinberg, Phys. Rev. Lett. **19**(21), 1264 (1967)
16. T. Aoyama, M. Hayakawa, T. Kinoshita, M. Nio, Phys. Rev. Lett. **99**(11), 110406 (2007)
17. D. Hanneke, S. Fogwell, G. Gabrielse, Phys. Rev. Lett. **100**(12), 120801 (2008)
18. L. Morel, Z. Yao, P. Cladé, S. Guellati-Khélifa, Nature **588**(7836), 61 (2020)
19. R.H. Parker, C. Yu, W. Zhong, B. Estey, H. Müller, Science **360**(6385), 191 (2018)
20. J.A. Frieman, M.S. Turner, D. Huterer, Annu. Rev. Astron. Astrophys. **46**, 385 (2008)
21. G. Bertone, D. Hooper, J. Silk, Phys. Rep. **405**(5–6), 279 (2005)
22. P.D. Group, et al., Chin. Phys. C **40**(10), 100001 (2016)
23. A.D. Sakharov, in *In The Intermissions... Collected Works on Research into the Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16* (World Scientific, 1998), pp. 84–87
24. G. Lüders, Ann. Phys. **2**(1), 1 (1957)
25. R. Jost, Acta **30**, 409 (1957)
26. D. Colladay, V.A. Kostelecký, Phys. Rev. D **55**(11), 6760 (1997)
27. R. Bluhm, in *Special Relativity* (Springer, Berlin, 2006), pp. 191–226
28. V.A. Kostelecký, S. Samuel, Phys. Rev. D **39**(2), 683 (1989)
29. M. Ahmadi, B. Alves, C. Baker, W. Bertsche, E. Butler, A. Capra, C. Carruth, C. Cesar, M. Charlton, S. Cohen et al., Nature **548**(7665), 66 (2017)
30. E. Widmann, M. Diermaier, B. Juhász, C. Malbrunot, O. Massiczek, C. Sauerzopf, K. Suzuki, B. Wünschek, J. Zmeskal, S. Federmann et al., Hyperfine Interact. **215**(1), 1 (2013)
31. S. Maury, W. Oelert, W. Bartmann, P. Belochitskii, H. Breuker, F. Butin, C. Carli, T. Eriksson, S. Pasinelli, G. Tranquille, Hyperfine Interact. **229**(1), 105 (2014)
32. M. Doser, et al., in *J. Phys.: Conf. Ser.*, vol. 199 (IOP Publishing, 2010), p. 012009
33. M. Fujiwara, G. Andresen, W. Bertsche, P. Bowe, C. Bray, E. Butler, C. Cesar, S. Chapman, M. Charlton, J. Fajans, et al., in *AIP Conference Proceedings*, vol. 1037 (American Institute of Physics, 2008), pp. 208–220
34. E. Widmann, C. Amsler, S.A. Cuendis, H. Breuker, M. Diermaier, P. Dupré, C. Evans, M. Fleck, A. Gligorova, H. Higaki et al., Hyperfine Interact. **240**(1), 1 (2019)
35. P. Perez, Y. Sacquin, Class. Quantum Grav. **29**(18), 184008 (2012)
36. T. Aumann, J. Grenard, S. Naimi, D. Rossi, Y. Ono, H. De Gersem, S. Sels, E. Siesling, A. Schmidt, D. Calvet et al., Puma: antiprotons and radioactive nuclei. Technical Report (2019)
37. N. Kuroda, S. Ulmer, D. Murtagh, S. Van Gorp, Y. Nagata, M. Diermaier, S. Federmann, M. Leali, C. Malbrunot, V. Mascagna et al., Nat. Commun. **5**(1), 1 (2014)
38. M. Diermaier, C. Jepsen, B. Kolbinger, C. Malbrunot, O. Massiczek, C. Sauerzopf, M. Simon, J. Zmeskal, E. Widmann, Nat. Commun. **8**(1), 1 (2017)
39. M. Hori, H. Aghai-Khozani, A. Sótér, D. Barna, A. Dax, R. Hayano, T. Kobayashi, Y. Murakami, K. Todoroki, H. Yamada et al., Science **354**(6312), 610 (2016)
40. S. Alighanbari, G. Giri, F.L. Constantin, V. Korobov, S. Schiller, Nature **581**(7807), 152 (2020)
41. F. Heiße, F. Köhler-Langes, S. Rau, J. Hou, S. Junck, A. Kracke, A. Mooser, W. Quint, S. Ulmer, G. Werth et al., Phys. Rev. Lett. **119**(3), 033001 (2017)

42. G. Andresen, M. Ashkezari, M. Baquero-Ruiz, W. Bertsche, P.D. Bowe, E. Butler, C. Cesar, S. Chapman, M. Charlton, A. Deller et al., Nature **468**(7324), 673 (2010)
43. M. Ahmadi, B. Alves, C. Baker, W. Bertsche, E. Butler, A. Capra, C. Carruth, C. Cesar, M. Charlton, S. Cohen et al., Nat. Commun. **8**(1), 1 (2017)
44. M. Ahmadi, M. Baquero-Ruiz, W. Bertsche, E. Butler, A. Capra, C. Carruth, C. Cesar, M. Charlton, A. Charman, S. Eriksson et al., Nature **529**(7586), 373 (2016)
45. A. Charman, Nat. Commun. **4**(1), 1 (2013)
46. M. Ahmadi, B.X.R. Alves, C.J. Baker, W. Bertsche, A. Capra, C. Carruth, C.L. Cesar, M. Charlton, S. Cohen, R. Collister et al., Nature **557**(7703), 71 (2018)
47. C. Baker, W. Bertsche, A. Capra, C. Carruth, C. Cesar, M. Charlton, A. Christensen, R. Collister, A.C. Mathad, S. Eriksson et al., Nature **592**(7852), 35 (2021)
48. C.G. Parthey, A. Matveev, J. Alnis, B. Bernhardt, A. Beyer, R. Holzwarth, A. Maistrou, R. Pohl, K. Predehl, T. Udem et al., Phys. Rev. Lett. **107**(20), 203001 (2011). https://doi.org/10.1103/PhysRevLett.107.203001
49. V.A. Kosteleckỳ, A.J. Vargas, Phys. Rev. D **92**(5), 056002 (2015)
50. B. Schwingenheuer, R.A. Briere, A.R. Barker, E. Cheu, L.K. Gibbons, D.A. Harris, G. Makoff, K.S. McFarland, A. Roodman, Y.W. Wah et al., Phys. Rev. Lett. **74**(22), 4376 (1995)
51. R.S. Van Dyck Jr, P.B. Schwinberg, H.G. Dehmelt, Phys. Rev. Lett. **59**(1), 26 (1987)
52. B. Abi, T. Albahri, S. Al-Kilani, D. Allspach, L. Alonzi, A. Anastasi, A. Anisenkov, F. Azfar, K. Badgley, S. Baeßler et al., Phys. Rev. Lett. **126**(14), 141801 (2021)
53. S. Borsanyi, Z. Fodor, J. Guenther, K.H. Kampert, S.D. Katz, T. Kawanai, T.G. Kovacs, S.W. Mages, A. Pasztor, F. Pittler, J. Redondo, A. Ringwald, K.K. Szabo, Nature **539**(7627), 69 (2016). https://doi.org/10.1038/nature20115
54. A. Keshavarzi, K.S. Khaw, T. Yoshioka, (2021). arXiv:2106.06723
55. G. Gabrielse, A. Khabbaz, D.S. Hall, C. Heimann, H. Kalinowsky, W. Jhe, Phys. Rev. Lett. **82**(16), 3198 (1999)
56. L.S. Brown, G. Gabrielse, Rev. Modern Phys. **58**(1), 233 (1986)
57. S. Ulmer, H. Kracke, K. Blaum, S. Kreim, A. Mooser, W. Quint, C.C. Rodegheri, J. Walz, Rev. Sci. Inst. **80**(12), 123302 (2009)
58. H. Nagahama, G. Schneider, A. Mooser, C. Smorra, S. Sellner, J. Harrington, T. Higuchi, M. Borchert, T. Tanaka, M. Besirli, K. Blaum, Y. Matsuda, C. Ospelkaus, W. Quint, J. Walz, Y. Yamazaki, S. Ulmer, Rev. Scie. Inst. **87**(11), 113305 (2016)
59. C. Smorra, A. Mooser, M. Besirli, M. Bohman, M.J. Borchert, J. Harrington, T. Higuchi, H. Nagahama, G.L. Schneider, S. Sellner, T. Tanaka, K. Blaum, Y. Matsuda, C. Ospelkaus, W. Quint, J. Walz, Y. Yamazaki, S. Ulmer, Phys. Lett., Sect. B: Nucl. Element. Particle High-Energy. Phys. **769**, 1 (2017). https://doi.org/10.1016/j.physletb.2017.03.024
60. S. Sturm, F. Köhler, J. Zatorski, A. Wagner, Z. Harman, G. Werth, W. Quint, C.H. Keitel, K. Blaum, Nature **506**(7489), 467 (2014)
61. C.d.C. Rodegheri, K. Blaum, H. Kracke, S. Kreim, A. Mooser, W. Quint, S. Ulmer, J. Walz, New J. Phys. **14**(6), 063011 (2012)
62. S. Ulmer, C.C. Rodegheri, K. Blaum, H. Kracke, A. Mooser, W. Quint, J. Walz, Phys. Rev. Lett. **106**(25), 253001 (2011)
63. A. Mooser, S. Ulmer, K. Blaum, K. Franke, H. Kracke, C. Leiteritz, W. Quint, C.C. Rodegheri, C. Smorra, J. Walz, Nature **509**(7502), 596 (2014)
64. S. Sellner, M. Besirli, M. Bohman, M.J. Borchert, J. Harrington, T. Higuchi, A. Mooser, H. Nagahama, G. Schneider, C. Smorra et al., New J. Phys. **19**(8), 083023 (2017)
65. C. Smorra, A. Mooser, K. Franke, H. Nagahama, G. Schneider, T. Higuchi, S.V. Gorp, K. Blaum, Y. Matsuda, W. Quint et al., Int. J. Mass Spectrom. **389**, 10 (2015)
66. K. Abe, Y. Haga, Y. Hayato, M. Ikeda, K. Iyogi, J. Kameda, Y. Kishimoto, M. Miura, S. Moriyama, M. Nakahata et al., Phys. Rev. D **95**(1), 012004 (2017)
67. S. Geer, J. Marriner, M. Martens, R. Ray, J. Streets, W. Wester, M. Hu, G. Snow, T. Armstrong, C. Buchanan et al., Phys. Rev. Lett. **84**(4), 590 (2000)
68. K. Babu, J.C. Pati, F. Wilczek, Phys. Lett. B **423**(3–4), 337 (1998)
69. R.J. Hughes, Phys. Rev. D **41**(8), 2367 (1990)

70. R.J. Hughes, M.H. Holzscheiter, Phys. Rev. Lett. **66**(7), 854 (1991)
71. O.W. Greenberg, Phys. Rev. Lett. **89**(23), 231602 (2002)
72. G. Chardin, G. Manfredi, Hyperfine Interact. **239**(1), 1 (2018)
73. M. Charlton, S. Eriksson, G.M. Shore, *Antihydrogen and Fundamental Physics* (Springer, Berlin, 2020)
74. J.K. Thompson, S. Rainville, D.E. Pritchard, Nature **430**(6995), 58 (2004)
75. K.R. Lykke, K.K. Murray, W.C. Lineberger, Phys. Rev. A **43**(11), 6104 (1991)
76. S. Rainville, J.K. Thompson, E.G. Myers, J.M. Brown, M.S. Dewey, E.G. Kessler, R.D. Deslattes, H.G. Börner, M. Jentschel, P. Mutti et al., Nature **438**(7071), 1096 (2005)
77. M. Charlton, S. Eriksson, G. Shore, Testing fundamental physics in antihydrogen experiments (2020). arXiv:2002.09348
78. I. Kenyon, Phys. Lett. B **237**(2), 274 (1990)
79. C. Tchernin, E.T. Lau, S. Stapelberg, D. Hug, M. Bartelmann, Astron. Astrophys. **644**, A126 (2020)
80. W.A. Bertsche, Philosoph. Trans. R. Soc. A: Math. Phys. Eng. Sci. **376**(2116), 20170265 (2018)
81. P. Scampoli, J. Storey, Modern Phys. Lett. A **29**(17), 1430017 (2014)
82. R. Bluhm, V.A. Kostelecký, N. Russell, Phys. Rev. D **57**(7), 3932 (1998)
83. Y. Ding, M.F. Rawnak, Phys. Rev. D **102**(5), 056009 (2020)
84. M. Borchert, P. Blessing, J. Devlin, J. Harrington, T. Higuchi, J. Morgner, C. Smorra, E. Wursten, M. Bohman, M. Wiesinger et al., Phys. Rev. Lett. **122**(4), 043201 (2019)
85. H. Nagahama, C. Smorra, S. Sellner, J. Harrington, T. Higuchi, M. Borchert, T. Tanaka, M. Besirli, A. Mooser, G. Schneider et al., Nat. Commun. **8**(1), 1 (2017)
86. H. Häffner, T. Beier, S. Djekić, N. Hermanspahn, H.J. Kluge, W. Quint, S. Stahl, J. Verdú, T. Valenzuela, G. Werth, Europ. Phys. J.D-Atomic. Mol. Opt. Plasma Phys. **22**(2), 163 (2003)
87. J. DiSciacca, M. Marshall, K. Marable, G. Gabrielse, S. Ettenauer, E. Tardiff, R. Kalra, D.W. Fitzakerley, M.C. George, E.A. Hessels et al., Phys. Rev. Lett. **110**(13), 130801 (2013)
88. P.F. Winkler, D. Kleppner, T. Myint, F.G. Walther, Phys. Rev. A **5**(1), 83 (1972)
89. Y. Stadnik, B. Roberts, V. Flambaum, Phys. Rev. D **90**(4), 045035 (2014)
90. Y.V. Stadnik, V.V. Flambaum, (2018). arXiv:1806.03115
91. C. Abel, N.J. Ayres, G. Ban, G. Bison, K. Bodek, V. Bondar, M. Daum, M. Fairbairn, V.V. Flambaum, P. Geltenbort et al., Phys. Rev. X **7**(4), 041034 (2017)
92. Y. Ding, V.A. Kostelecký, Phys. Rev. D **94**(5), 056008 (2016)
93. P. Sikivie, Phys. Rev. Lett. **51**(16), 1415 (1983)
94. I.G. Irastorza, J. Redondo, Progr. Particle Nucl. Phys. **102**, 89 (2018). https://doi.org/10.1016/j.ppnp.2018.05.003
95. P. Sikivie, N. Sullivan, D.B. Tanner, Phys. Rev. Lett. **112**(13), 131301 (2014)
96. S. Ulmer, K. Blaum, H. Kracke, A. Mooser, W. Quint, C.C. Rodegheri, J. Walz, Nucl. Inst. Methods Phys. Res. Sect. A **705**, 55 (2013). https://doi.org/10.1016/j.nima.2012.12.071
97. M. Bohman, A. Mooser, G. Schneider, N. Schön, M. Wiesinger, J. Harrington, T. Higuchi, H. Nagahama, C. Smorra, S. Sellner et al., J. Modern Optics **65**(5–6), 568 (2018)
98. M. Bohman, V. Grunhofer, C. Smorra, M. Wiesinger, C. Will, M. Borchert, J. Devlin, S. Erlewein, M. Fleck, S. Gavranovic et al., Nature **596**(7873), 514 (2021)
99. C. Will, M. Bohman, T. Driscoll, M. Wiesinger, F. Abbass, M. Borchert, J. Devlin, S. Erlewein, M. Fleck, B. Latacz et al. (2021). arXiv:2112.04818
100. J.M. Cornejo, R. Lehnert, M. Niemann, J. Mielke, T. Meiners, A. Bautista-Salvador, M. Schulte, D. Nitzschke, M.J. Borchert, K. Hammerer et al. (2021). arXiv:2106.06252

NA64—Search for Dark Sector Particles

Dipanwita Banerjee

1 Introduction

One of the main triumphs in the field of science and knowledge was a theory, developed in stages, through the works of various scientists along the latter half of the 20th century that explains nearly everything that rule our daily lives—from the fundamental structure of matter and energy to reactions that power the sun, from understanding the first moments of our universe's existence to the building blocks of our being. Physicist Robert Oerter called it "The Theory of Almost Everything" in his book named the same [1], but it is more commonly called the "Standard Model of Elementary Particles". The Standard Model consists of the fundamental particles called the fermions that include the quarks and leptons. It also includes the gauge bosons which are the force carriers as well as the Higgs boson that is responsible for the origin of mass [2]. Even though the Standard Model has been successful in explaining almost everything, it falls short in being a complete theory of fundamental forces and fails to explain some phenomena. The Dark Matter, for example is not part of the Standard Model and points to physics "Beyond the Standard Model". One of the hypothesis for Dark Matter points to an existence of a dark sector with its own particles and forces similar to the Standard Model as shown in Fig. 1. This sector exists in parallel with our "normal" sector and will not interact with it via known forces except via gravity. However, there can be some additional interactions via some portals (Review available in [3]). Analogous to electromagnetism of our world, for which the massless gauge boson, the photon, transmits force between charged particles, there could be a dark electromagnetism with a possibly massive gauge boson, called the dark photon that transmits the forces between dark particles. Our photon can interact with a dark photon within a certain mass range and coupling strength resulting in a new additional interaction between our and the dark sector. In

D. Banerjee (✉)
CERN, Espl. des Particules 1, 1211 Meyrin, Geneva, Switzerland
e-mail: dipanwita.banerjee@cern.ch

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_11

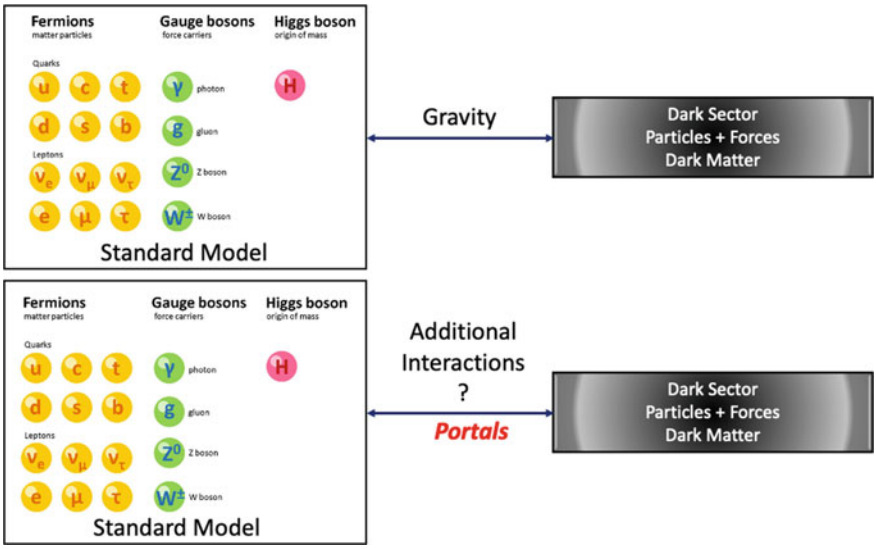

Fig. 1 Standard model of elementary particles and the dark sector including dark matter interacting via gravity as well as some additional interactions via portals

the most interesting class of models the A' is light with sub-GeV masses and have a $\gamma - A'$ coupling strength, $\epsilon << 1$ and in the range which can be experimentally tested.

An additional motivation for new dark particle searches comes from the observed discrepancy between the predicted and the experimental values of the anomalous magnetic moment, $(g-2)_\mu$ of the muon [4]. All elementary particles are like a spinning top and thus has an intrinsic angular momentum called the spin. When a charged particle spins, including the muons, it will generate their own magnetic field. The strength of this magnetic field is referred to as the magnetic moment or the g-factor. Any deviation from the theoretical prediction of this g will point to new physics. Almost two decades ago the Brookhaven National Laboratory presented their first results where a 3.6σ deviation was observed in the $(g-2)_\mu$ [5]. In 2021 the results were confirmed by Fermilab with a 4.2σ discrepancy [6]. Calculations have shown that this could be possible if a sub-GeV dark boson, Z_μ, which interacts presumably only with muons (and tau leptons) with a coupling strength, $\epsilon \sim 10^{-3}$, exists (as the A' explanation for the $(g-2)_\mu$ anomaly has since been excluded). Similar to the dark photon, the Z_μ could also induce additional interaction between the ordinary and dark sectors not predicted by the Standard Model. That model is very attractive as it could explain both the $(g-2)_\mu$ anomaly and the Dark Matter puzzle (see Sect. 4)

The most effective search for the A' is through its decays. Depending on the mass of the A' it can decay either visibly to Standard Model leptons, $l = e$, μ or hadrons when the $M_{A'} > 2M_l$ where $M_{A'}$ is the mass of the dark photon and M_l is the mass

of the leptons. The A' could also decay invisibly to dark states with masses $< M_{A'}/2$. When NA64 started its run the $(g-2)_\mu$ explanation for A' decaying visibly only had been excluded by previous beam dump, fixed target, collider and rare meson decay experiments [7]. NA64 therefore started with focussing on the $A' \rightarrow invisible$ decay search.

2 NA64 Experiment

NA64 is an experiment at the CERN SPS approved in March 2016 with the goal to search for $A' \rightarrow$ invisible decay with an electron beam [8]. A dark photon can be produced in any reaction that can produce a normal photon. When an electron beam hits a target material it produces bremsstrahlung photons. The dark photons can then be produced via its mixing with the normal photon. NA64 aims to exploit this interaction to search for the dark photons.

2.1 Principle of the Experiment

NA64 is a fixed target experiment that combines the beam dump technique with missing energy measurement searching for invisible decays of massive A' produced in the reaction:

$$e^- Z \rightarrow e^- Z A' \tag{1}$$

It employs the 100 GeV electron beam from the CERN SPS H4 beamline. The beam is dumped on the active target, the ECAL, as shown in Fig. 2. The beam with energy $E_0 = 100$ GeV interacts in the target and deposits an energy, E_{Ecal}, in the ECAL. The A' if produced carries the remaining energy $E_0 - E_{Ecal}$. After escaping the target the A' then decays into dark matter particles. As the dark matter particles will not interact with the downstream detectors and energy should be conserved the undetected energy of $E_0 - E_{Ecal}$ above a certain threshold gives the signal of the dark photon. In the situation when A' is not produced the remaining energy is carried by Standard Model particles which are detected in the downstream part of the NA64 setup as E_{Hcal}. The threshold for the missing energy set by NA64 is 50% of the incident energy so a missing energy >50 GeV indicates a positive signal for the A' search. The complete setup of NA64 is shown in Fig. 3. As NA64 is looking for missing energy there are three main sources of errors that should be minimised.

1. The incoming energy of the beam itself is <50 GeV → If the incoming energy of the beam is <50 GeV and the beam deposits all its energy in the ECAL, E_{Hcal} will be zero and will mimic a missing energy signal.

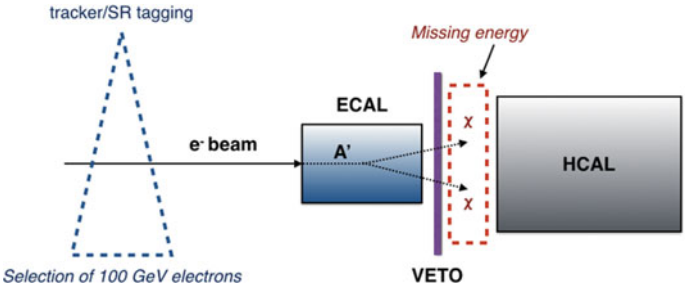

Fig. 2 Schematic for the NA64 invisible decay search

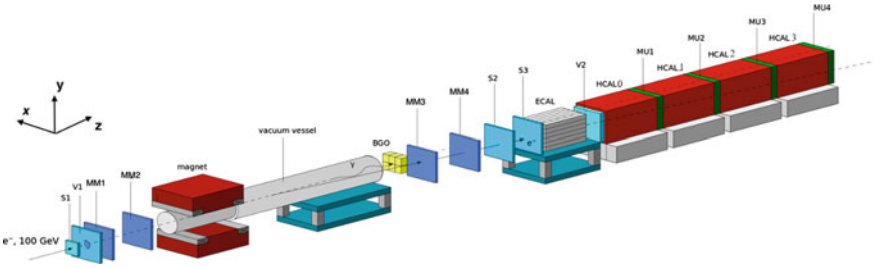

Fig. 3 Full setup of NA64 for the invisible search

2. The incoming beam is not an electron → If the incoming beam has any contamination of hadrons which are not identified this can also create background for the search.
3. There is an energy leak from the downstream part → If the NA64 setup does not detect part of the energy downstream it can mimic a missing energy signal.

The methods used to suppress the backgrounds are discussed in the following subsections.

2.1.1 Suppression of Low Energy Beam Electrons

In order to prevent misidentification of energy of the incoming beam NA64 comprises of a tracking section in its upstream part that includes the MM1-4, which are the tracking stations giving the hit positions of the beam on these planes, and the magnet as shown in Fig. 3. Any charged particle in a magnetic field moves in a circular motion with the radius depending on its momentum for a given magnetic field as:

$$p = qBr \tag{2}$$

where p is the momentum of the beam, q is the charge, B is the magnetic field and r is the radius of curvature. Therefore for a known B and the calculated r from the hit

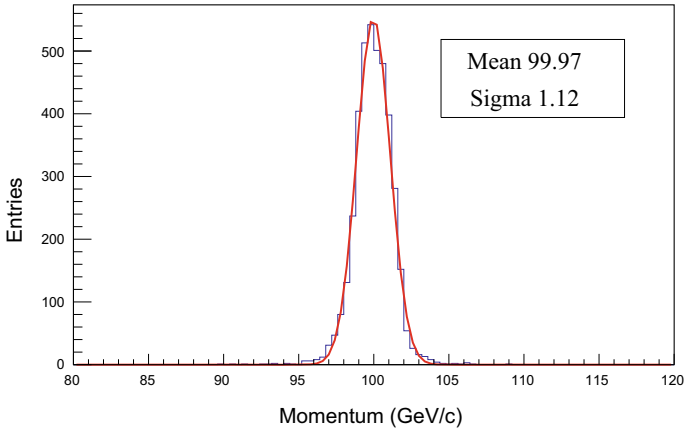

Fig. 4 Momentum resolution for the reconstructed momentum in NA64

positions of the particle in the tracking stations one can measure the momentum of the incoming beam. NA64 uses a 7 T.m dipole magnet for the momentum reconstruction and tracking detectors with a spatial resolution ∼100 μm [9]. The resolution of the reconstructed momentum is ∼1% for a 100 GeV beam as shown in Fig. 4.

2.1.2 Suppression of Hadron Contamination

A charged particle in a magnetic field moves in a circular motion emitting photons along its trajectory called the synchrotron radiation. The total power, S, emitted per unit length by a relativistic charged particle of energy E with mass M and with bending radius r in a magnetic field B perpendicular to its velocity is given by:

$$S = \frac{q^2 c}{6\pi} \frac{1}{(Mc^2)^4} \frac{E^4}{R^2} \tag{3}$$

where q is the charge of the particle and c is the speed of light. The synchrotron photons are emitted tangentially to the particle trajectory. As seen from the Eq. 3 the total emitted power scales inversely with the fourth power of the charged particle mass. Therefore, the synchrotron radiation emitted by heavy charged particles is orders of magnitude less than the light ones and can be used to suppress them. Heavy charged particles like $\mu^{+/-}$ and $\pi^{+/-}$ which has about 200 e^- mass, radiate $\sim 10^9$ times less than an electron. However this is true for ideal vacuum, which is unlike real experimental setup where interaction of hadrons in vacuum windows, residual gas, instrumentations etc., limit the suppression factor due to emission of secondary particles with enough kinetic energy (several MeVs) that mimics the synchrotron radiation of an electron. This was checked during the beam time and validated with

Monte Carlo simulation. The suppression factor thus achieved in NA64 $< 10^{-3}$ for the hadrons with a beam contamination of $\pi/e < 10^{-2}$ [10].

2.1.3 A Fully Hermetic Detector

The $A' \rightarrow$ invisible search requires not only a precise knowledge of the incoming beam but also an accurate measurement of the missing energy from the incoming beam's interaction. The electromagnetic calorimeter is crucial to determine precisely the deposited energy by the incoming electron in the active target and hence calculate the missing energy downstream. An electromagnetic calorimeter, the ECAL, with $\sim$40 radiation lengths (X_0) is used in NA64. A radiation length is the characteristic distance traversed by an electron before it loses its energy by a factor of $\frac{1}{e}$ via electromagnetic interaction. Therefore the 40 radiation lengths ensure that all the energy from electromagnetic interaction will be deposited within the ECAL for a precise measurement of the energy. As discussed the knowledge of E_{Ecal} is extremely important to estimate the missing energy.

The longitudinal hermeticity of the experimental setup is enhanced by using hadronic calorimeters, the HCAL modules, to detect charged and neutral hadrons and subsequently provide a measurement of the missing energy, not detected in the ECAL, after proper tagging of the incoming particle with the trackers and synchrotron radiation detectors. The HCAL consists of four modules, each comprising of 7 interaction lengths. An interaction length is the mean distance that a hadronic particle covers before undergoing an inelastic nuclear interaction. Therefore a total of 28 interaction lengths with the four HCAL modules ensure that any hadron that escapes the ECAL will interact in them and be absorbed which will give a precise measurement of the energy, E_{Hcal}.

Under Standard Model interactions:

$$E_{Ecal} + E_{Hcal} = E_0 \tag{4}$$

where E_0 is the energy of the incoming particle and for the dark photon signal:

$$E_{Ecal} + E_{Hcal} < 0.5E_0 \tag{5}$$

Since the signal is characterized by "No Interaction" in the HCAL the zero energy threshold had to be checked for the detectors as well. In order to estimate the zero energy event threshold Monte Carlo simulations were performed and compared with data for a 100 GeV electron beam. The threshold for the HCAL for a zero energy event was then set at 1 GeV i.e., if $E_{Hcal} < 1$ GeV it is taken as a zero energy event.

With these selection criteria the data collected by NA64 have been analysed and the current results from the invisible decay search are detailed in the following Section.

3 Current Results of the NA64 A' → Invisible Search Experiment

Since 2016 NA64 has collected $\sim 3 \times 10^{11}$ electrons on target until 2018. The analysis of the 2016 data has been completed and that for the 2017–2018 data is currently being finalised. The missing energy signal has been searched with the selection

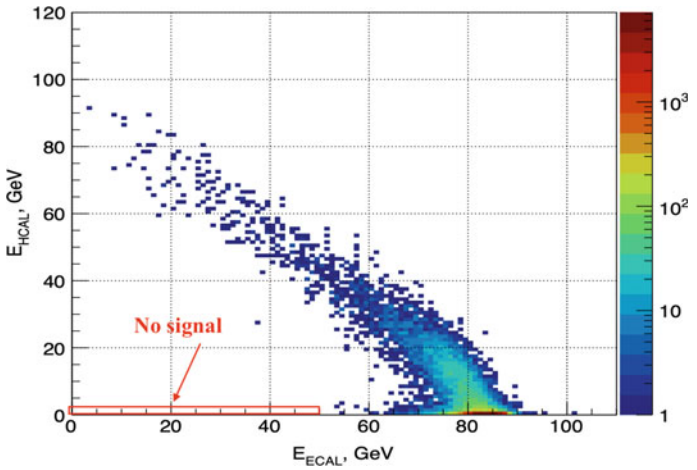

Fig. 5 The energy in the HCAL as a function of the energy in the ECAL as detected in NA64. The red box corresponds to the signal region for A' which shows no event

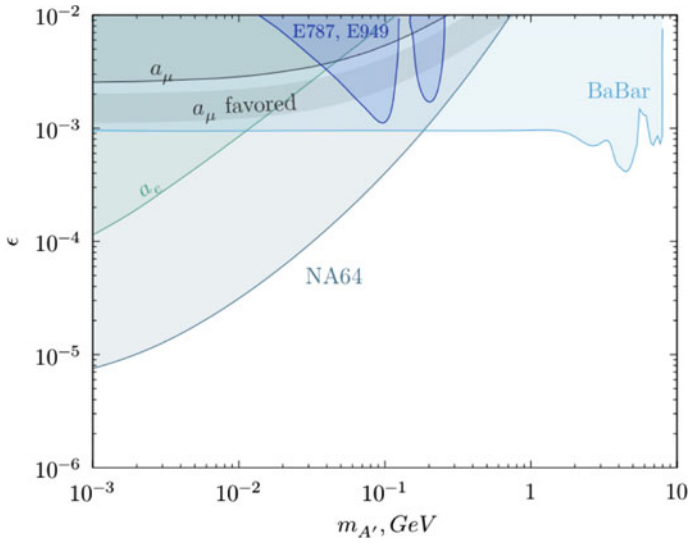

Fig. 6 The coupling strength, ϵ, and mass, $M_{A'}$, parameter space of the A' including the region excluded by NA64 with full statistics collected between 2016 and 2018

criteria described above. Figure 5 shows the 2-dimensional plot of the energy in the HCAL as a function of that in the ECAL. The red box corresponds to the signal region where the $E_{Ecal} < 0.5E_0$ and E_{Hcal} corresponds to no signal i.e., $\leq$1 GeV. As no event has been detected in the missing energy measurement NA64 was able to exclude a considerable region of the coupling strength—mass parameter space of the A' as shown in Fig. 6 [11]. This also includes the A' explanation of $(g-2)_\mu$ which thus could be excluded for the invisible decay channel. With more statistics NA64 would be able to put very competitive limits on the parameter space for the sub-GeV dark matter or potentially discover the dark photon.

4 Additional Searches for Dark Particles

Although NA64 was primarily designed for the $A' \rightarrow$ invisible search it can also be used for other complementary searches. As discussed above A' could also decay visibly into Standard Model leptons depending on its mass. NA64 could also search for these visible decays with a small modification to its setup as shown in Fig. 7. In this case a 100 GeV electron beam is incident in a similar way on an active target, WCAL. If the A' is produced it will escape the WCAL and will be undetected in the downstream detector the V2. Then it will decay into the e^+e^- pair which can be detected with the trackers downstream. Therefore, the visible search can be equated to a "light shining through a wall" signature. During the 2017 and 2018 runs NA64 has collected upto 8×10^{10} electrons on target in this mode. The data has been analysed and no event has been detected in the signal region. NA64 was thus able to exclude part of the parameter space for the visibly decaying dark photon as well as shown in Fig. 8 [12].

In addition NA64 also plans to search for the dark sector through particles that predominantly couple to muons and taus, the heavier counterparts of the electrons with its experiment called the NA64μ [13]. As mentioned the A' explanation for the $(g-2)_\mu$ anomaly has already been excluded, however dark sector models suggest that there can be new massive gauge boson, Z_μ which couples predominantly to μ and τ and can explain this discrepancy [14]. This is an additional motivation for such searches along-with dark matter. Such particles can be searched for in a μ

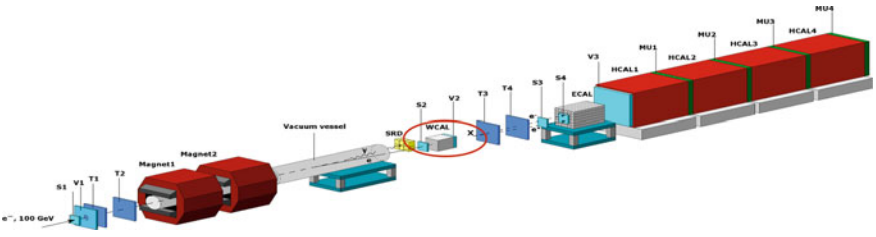

Fig. 7 Full setup of NA64 for the $A' \rightarrow$ visible search

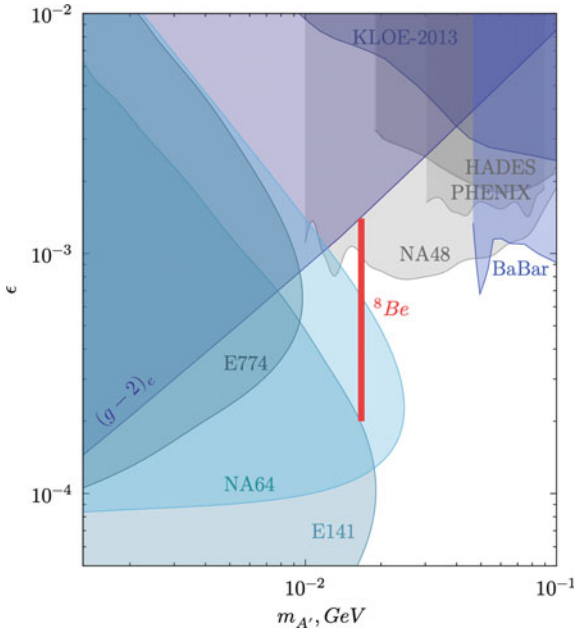

Fig. 8 The parameter space excluded by NA64 for the $A' \rightarrow$ visible channel using the 2017 data (blue shadowed region) and the 2017+2018 data (dashed line, preliminary)

interaction for example. NA64 therefore plans to extend its electron beam search with muon beams using the M2 beamline at the CERN SPS.

The M2 beamline at the CERN SPS delivers high intensity high energy muon beams [15]. NA64 plans to exploit this beamline to search for $Z_\mu \rightarrow$ invisible decays. The search mechanism is slightly different compared with the electron beam as muons are much heavier than electrons and therefore does not get stopped in the ECAL. For this search a 160 GeV muon beam is momentum selected with the NA64 spectrometer similar to the electron beam described above. The muon beam is then incident on the ECAL where the beam interacts to potentially produce the Z_μ via interaction of the muon on the target. In case of the Z_μ production it carries majority of the energy $>0.5E_0$ where E_0 is the incident muon beam energy, and subsequently decays into dark matter particles which remains undetected in the downstream part of the setup similar to the $A' \rightarrow$ invisible case. However, unlike the electrons the muons will not deposit its energy but will carry the remaining $<0.5E_0$ and escape the ECAL. Therefore there are downstream trackers with spectrometer magnets similar to the one for the incoming momentum reconstruction to reconstruct the momentum of the outgoing muon. This search is therefore not a missing energy search but a missing momentum search. The schematic of the setup is shown in Fig. 9. The signal includes that the reconstructed momentum of the outgoing muon <80 GeV/c and there is no energy in the ECAL and HCAL while the reconstructed momentum of the incoming muon is $\sim$160 GeV/c. In 2021 November NA64 had its first pilot run for the muon beam in the M2 beamline and accumulated $\sim 5 \times 10^9$ μ on target. The first aimed

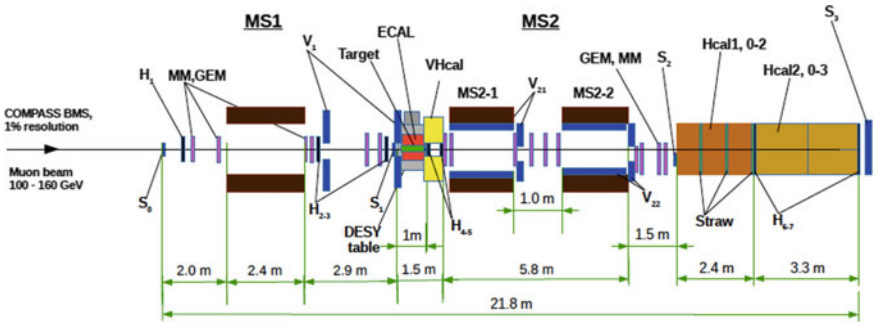

Fig. 9 Full setup of NA64 for the $Z_\mu \rightarrow$ invisible search

milestone for NA64μ is to collect 10^{10} μ on target to be able to cover the $(g-2)_\mu$ favoured parameter space. In 2022 2–3 weeks of beam time have been requested to potentially reach this goal.

5 Summary

Dark Sector physics can be effectively probed with the NA64 experiment and its search techniques in the medium-term future. NA64 uses two complementary approaches including beam-dump and missing energy/momentum and can either discover or rule-out nearly all predictive models of sub-GeV thermal Dark Matter. By the Long Shutdown 3 in 2025 NA64 aims to collect $\sim 5 \times 10^{12}$ electrons on target and with the improved statistics NA64 can probe most of the existing thermal Dark Matter models [16]. With the restart of the CERN accelerator complex in 2021 and the planned runs the new results are expected to be extremely promising and exciting.

References

1. R. Oerter, *The Theory of Almost Everything: The Standard Model, the Unsung Triumph of Modern Physics* (2006). ISBN 0-13-236678-9
2. M.K. Gaillard, P.D. Grannis, F.J. Sciulli, The standard model of particle physics. Rev. Mod. Phys **71**, S96 (1999)
3. J. Jaeckel, A. Ringwald, The low-energy frontier of particle physics. Ann. Rev. Nucl. Part. Sci. **60**, 405 (2010)
4. J.P. Miller, E. de Rafael, B. Lee Roberts, Muon (g-2): experiment and theory. Rep. Prog. Phys. **70**, 795 (2007)
5. G.W. Bennett, B. Bousquet et al., Measurement of the negative muon anomalous magnetic momentum to 0.7 ppm. Phys. Rev. Lett. **92**, 161802 (2004)

6. First Results from Fermilab's Muon g-2 experiment strengthen evidence of new physics (2021). https://news.fnal.gov/2021/04/first-results-from-fermilabs-muon-g-2-experiment-strengthen-evidence-of-new-physics/
7. J. Alexander et al., Dark Sector 2016 Workshop: Community Report (2016). arXiv:1608.08632
8. NA64 Collaboration: D. Banerjee et al., Search for invisible decays of sub-GeV dark photons in missing energy events at the CERN SPS. Phys. Rev. Lett. **118**, 011802 (2017)
9. NA64 Collaboration: D. Banerjee et al., Performance of multiplexed XY resistive micromegas detectors in a high intensity beam. Nucl. Instrum. Meth. A **881**, 72–81 (2018)
10. NA64 Collaboration: E. Depero et al., High purity 100 GeV electron identification with synchrotron radiation. Nucl. Instrum. Meth. A **866**, 196 (2017)
11. M. Kirsanov, Recent results of the NA64 experiment at the CERN SPS. EPJ Web Conf. **212**, 06005 (2019)
12. NA64 Collaboration: D. Banerjee et al., Improved limits on a hypothetical X(16.7) boson and a dark photon decaying into e+e- pairs. Phys. Rev. D **101**(2020), 071101 (2020)
13. H. Sieber et al., Prospects in the search for a new light Z' boson with the NA64μ experiment at the CERN SPS (2021). arXiv:2110.15111
14. S.N. Gninenko, N.V. Krasnikov, Probing the muon gμ-2 anomaly, Lμ-Lτ gauge boson and Dark Matter in dark photon experiments. Phys. Lett. B **783**, 24–28 (2018)
15. N. Doble, L. Gatignon, G. von Holtey, F. Novoskoltsev, The upgraded Muon Beam at the SPS. CERN-SL 93-26 (EA) (1993)
16. S.N. Gninenko, D.V. Kirpichnikov, M.M. Kirsanov, N.V. Krasnikov, Combined search for light dark matter with electron and muon beams at NA64. Phys. Lett. B **796**, 117–122 (2019)

Space Searches

The AMS Experiment on the International Space Station

Maura Graziani and Nicola Tomassetti

1 Introduction

On the 16th of May 2011, NASA's Space Shuttle *Endeavour* was launched from Cape Canaveral at the Kennedy Space Center (Florida, USA) for its last mission. The main purpose of the mission was the delivery of one of the most complex instruments ever built: the Alpha Magnetic Spectrometer (AMS), a multipurpose detector of high-energy particles designed to operate in space for a long-duration mission of fundamental physics research. A few hours after the launch, a first power-on of AMS inside the shuttle was performed remotely, with a preliminary monitoring on ground of various temperatures and electrical currents. On May 19th, AMS was powered on for installation on the International Space Station. At 4:46 a.m. the same day the installation was completed and the first activation of the full experiment in space took place. Since then, AMS is sending us a continuous flow of scientific data, adding new knowledge on a variety of fundamental physics questions. The main AMS scientific goals include the indirect search of dark matter, the search of primordial antimatter, and the understanding of fundamental astrophysical processes of cosmic rays in the Galaxy. Cosmic rays are charged particles with galactic and extra-galactic origin, coming from outer space to the Earth's atmosphere with a broad range of energy that extends over several decades. Their origin, their transport in the Galaxy, and their interactions with the matter are not yet well known, so they are subject of extensive research. Cosmic particles are mainly composed by protons ($\sim$90%), helium ($\sim$8%), heavier nuclei ($\sim$1%), electrons ($\sim$1%) and antiparticles ($<$1%). Part of them such as electrons, protons, ^{4}He, C-N-O, or Fe are believed to be of primary origin, i.e., accelerated by explosions of supernova remnants or stellar

M. Graziani (✉) · N. Tomassetti
Department of Physics and Geology, University of Perugia, 06123 Perugia, Italy
e-mail: maura.graziani@unipg.it

N. Tomassetti
e-mail: nicola.tomassetti@unipg.it

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_12

winds, although the exact mechanisms are not yet well known [1]. Rarer CR elements such as ^{2}H, ^{3}He, Li-Be-B elements or antiparticles are believed to be of secondary origin, i.e. produced by collisions of primary nuclei with the interstellar medium. Measuring these components is crucial to understand the fundamental processes of CR acceleration and transport in the Galaxy. Moreover, the knowledge of the energy spectra of CR nuclei enables us to predict the level of secondary antimatter production in CR, which constitutes the astrophysical background for the search of dark matter induced signals. The energy spectra of CR antiparticles, such as antiprotons and positrons, are in fact recognized to be powerful observables for the search of dark matter signatures.

2 The AMS Experiment

AMS is a detector of high-energy charged particles operating onboard the International Space Station since May 2011 to conduct a long term observation of cosmic rays in near earth orbit. AMS has been conceived to address open questions in fundamental physics and astrophysics by means of accurate measurements of CR. The apparent absence of anti-matter in the Universe is still a mystery which questions the foundations of elementary particle physics and our current understanding of the origin and evolution of the cosmos. The detection of even a single anti-nucleus in the cosmic radiation, such a nucleus of $\bar{H}e$ or $\bar{C}$, could be a direct proof of the existence of antimatter domains, since the probability of their generation via spallation processes in the interstellar medium is very low [2] due to their high mass. Searches for nuclear anti-matter has been carried by now since more than 30 years with increasing sensitivity in continuously extending energy ranges, with no positive detection [3]. The large amount of CR particles that AMS will be able to collect during its mission (as of today more than 190 billion particles) will allow to search for anti-nuclei with unprecedented sensitivity.

Another primary discovery potential for AMS concerns the indirect dark matter detection. Dark matter could be detected in CR through its annihilation into standard particles and γ resulting in deviations or structures to be seen in the cosmic ray spectra. Due to the low intensity of fluxes coming from DM annihilation, detection of an excess can be possible only in the rarest CR components, i.e. anti-protons, positrons or anti-deuteron (never observed so far) which are not routinely produced at astrophysical sources.

A further objective of AMS, is the accurate measurements of all the charged species of the CR, including chemical species up to Iron and isotopes up to Carbon. These measurements are needed to address open questions in the origin and propagation of CR in our galaxy.

With a size of $2\times3\times4$ m^3 size for an overall weight of 7.5 tons and a total power consumption <2 kW, AMS takes full advantage of the state of-the art detector technology developed for high energy experiments operating at particle accelerators. I The instrument will be active for the entire ISS lifetime, i.e., until 2028 or beyond.

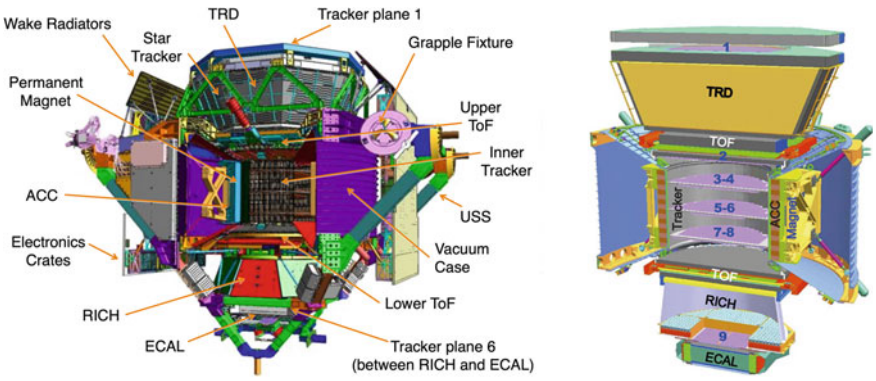

Fig. 1 Two schematic views of AMS. The various detectors are labelled

With this long observation time and its large acceptance ($\sim$0.5 m^2sr), AMS can provide high-quality data on CR fluxes at the TeV energy scale with unprecedented precision and sensitivity. Furthermore, the measurements of low energy CR fluxes over an entire 11-year solar cycle will enable us to perform a multichannel investigation of the solar modulation effect of Galactic CR.

2.1 The Detector

To fulfill its ambitious scientific objectives, AMS has been conceived as a particle physics experiment, with a high degree of redundancy in measuring the characteristics of the incident particles, satisfying at the same time the stringent requirements in reliability that are needed to survive to the hostile environment in space. Figure 1 shows a schematic view of the detector: the core of the instrument is the magnetic spectrometer composed of a permanent magnet, which produces a magnetic field with an intensity of 0.15 T, and of 9 layers of double-sided silicon micro-strip sensors that constitute the Tracker.

The task of the spectrometer is the measurement of particle rigidity ($R = P/eZ$, momentum/charge ratio) by mean of the reconstruction of the particle trajectory along the Tracker layers with $\sim 10\,\mu$m ($\sim 30\,\mu$m) of spatial resolution on the Y (X) side. Above and below the spectrometer two planes of time of flight counters (ToF) provide the main trigger of AMS and distinguish between up-going and down-going particles. This information, combined with the trajectory curvature given by the spectrometer, is used to reconstruct the sign of the charge. The spectrometer and the ToF constitute the key instruments for matter-antimatter separation. A Transition Radiation Detector (TRD) is located at the top of the instrument. The detector ends with a Ring Imaging Cherenkov detector (RICH) and an electromagnetic calorimeter (ECAL). The central part of AMS is surrounded by an anti-coincidence system (ACC)

that provides the veto signal in the trigger for the particles detected outside the field of view of the instrument.

The particle absolute charge (Z) is measured independently by each Tracker layers, by the ToF and the RICH. This redundancy allows to evaluate the fragmentation of high Z nuclei along the detector. The velocity $\beta = v/c$ can be determined from the transit time between the ToF planes (for Z = 1, $\Delta\beta/beta \sim 3\%$), or more precisely using the RICH system (for Z = 1, $\Delta\beta/beta \sim 0.1\%$).

In the measurement of CR leptons, the signals in TRD and ECAL are used to discriminate the leptonic component from the hadronic background. The combination of these information allows for an efficient lepton/hadron separation power. AMS instrument is described in details in Sect. 3.

3 AMS Results over 10 Years of Data Taking

In the following the main results obtained by AMS over 10 years of data taking will be discussed. The results include the fluxes of positrons, electrons, antiprotons, protons, and nuclei. The isotropic flux for a given specie x of CR, Φ_i, in the ith bin of energy ΔE_i, or in the ith bin of rigidity ΔR_i is given by:

$$\Phi_i = \frac{N_i}{A(1+\delta_i)T_i \Delta E_i (\Delta R_i)} \tag{1}$$

where:

- N_i is the number of events tagged as x in bin i corrected for the bin to-bin migration using the unfolding procedure described in [4].
- A_i is the corresponding effective acceptance that includes geometric acceptance, and the trigger and selection efficiencies, and is calculated from Monte Carlo simulation.
- T_i is the data collection time.
- δ_i is the data/MonteCarlo correction to the Acceptance, estimated by comparing the efficiencies in data and Monte Carlo simulation of every selection cut using information from the detectors unrelated to that cut.

3.1 Results on CR Leptons

Electrons (e^-) and positrons (e^+) constitute respectively the $\sim 1 - 0.1\%$ and $\sim 0.1 - 0.01\%$ (depending on energy) of the total CR radiation. The observed e^- are mainly Primaries, i.e. directly produced from the CR sources, while e^+ are secondaries, i.e. produced by the collisions of CR with the interstellar medium. Even if they constitute a rare component of CR, $e^{\pm}$ carry important physics information. Due to their low

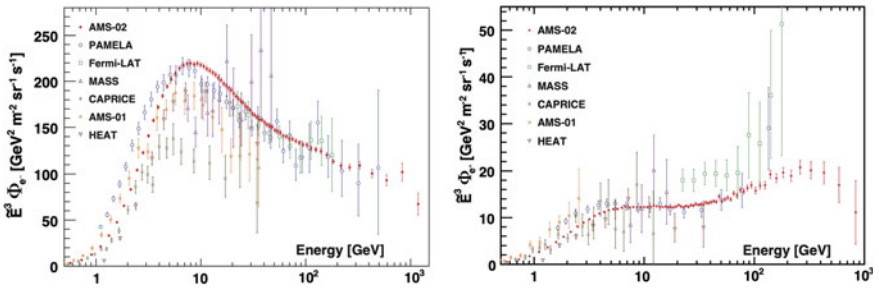

Fig. 2 The AMS electron (on the left) and positron (on the right) spectrum multiply by E^3 together with earlier measurements from PAMELA [5], Fermi-LAT [6], MASS [7], CAPRICE [8], AMS-01 [9], and HEAT [10, 11]

mass, are subject to important energy losses due to interaction with the interstellar medium, in their trajectory between the sources and Earth, representing a unique probes to study the CR sources property in the galactic neighborhood and a gold channel for the indirect search of dark matter.

The first publication on this subject from PAMELA, in early 2009, clearly showed a steady increase of the positron fraction [5] and this opened the road to a world-wide effort to account for new e^+ sources contributing to the flux. Measurements from Fermi were confirming the trend shown by PAMELA, even if the dispersion of the FERMI data with respect to the PAMELA and the following AMS measurements clearly shows the limits of the indirect charge measurement technique. The real breakthrough in these measurements came however with AMS results ams, ams2, ams3, ams4, ams5. Thanks to the large acceptance and the high electron/proton discriminating power, obtained combining the independent information coming from TRD and ECAL, AMS has provided precise measurements in an extended energy range.

Figure 2 shows the AMS electron (on the left) and positron (on the right) spectrum multiply by E^3 together with earlier experiments [5–10]. The electron flux has been measured in the energy range from 0.5 GeV to 1.4 TeV based on 28.1×10^6 electrons. The positron flux has been measured in the energy range from 0.5 GeV to 1 TeV and is based on 1.9 million of positrons. The AMS data significantly extend the measurements into the uncharted high-energy region. The behavior of the electron and positron as a function of energy is distinctly different.

The rise observed in the e^+ flux starting from 25.2±8 GeV, compared with the lower-energy trend, can be explained only taken into account an extra source of e^+ respect to the secondary production in the interstellar medium. As shown in Fig. 3 (left) in the entire energy range the e^+ flux is well described by the sum of two terms: "diffuse" term associated with the secondary positrons from cosmic ray collisions and the source term which has an exponential energy cutoff of 810^{+310}_{-180} GeV (see [12] for more details).

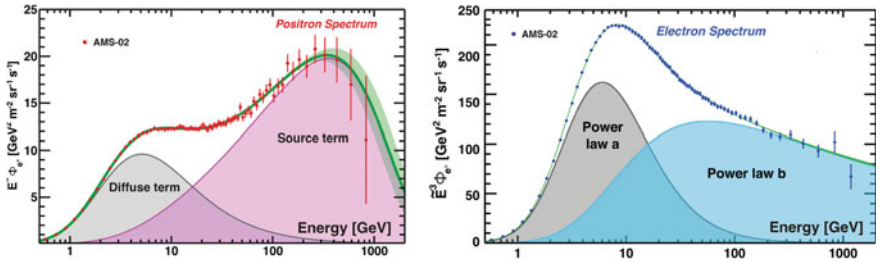

Fig. 3 On the left: The red data points represent the measured positron flux multiply by E^3. The source term contribution is represented by the magenta area and the diffuse term contribution by the gray area. On the right: the blue data points represent the measured electron flux multiply by E^3. The two power law components a and b are represented by the gray and blue areas, respectively

The source term can originate either from dark matter annihilation or from other astrophysical sources and the current AMS data are consistent whit both hypothesis [13–17]. More statistics at high energy is needed in order to assess the nature of the the source.

As shown in Fig. 3 (right) the electron flux is well described by the sum of two power law components. The flux exhibits a significant excess starting from $42.1^{+5.4}_{-5.2}$ GeV compared to the lower energy trends, but the nature of this excess is different from the positron flux excess above 25.2±8 GeV. Contrary to the positron flux, which has an exponential energy cutoff of 810^{+310}_{-180} GeV, at the 5σ level the electron flux does not have an energy cutoff below 1.9 TeV. The different behavior of the CR electrons and positrons measured by AMS is a clear evidence that most high energy electrons originate from different sources than high energy positrons.

An hint on the the nature of the source can be address from the measurement of anisotropies of electron and positron fluxes. Astrophysical point sources like pulsars will imprint a higher anisotropy on the arrival directions of energetic positrons [18, 19] than a smooth dark matter halo. If the excess of positrons has a dark matter origin, it should be isotropic. Both electron and positron fluxes are found to be consistent with isotropy with an upper limit on the amplitude of the dipole anisotropy for any axis in galactic coordinates of $\delta < 0.019$ (for positrons) and $\delta < 0.005$ for electrons at the 95% confident level [20].

3.2 Results on CR Nuclei

The first results on cosmic nuclei from AMS came in 2015 with the publication of the light nuclei flux of Hydrogen (protons) and Helium [4, 22]. The AMS data show that both species are described by a broken power-law in rigidity, as the fluxes experience a progressive spectral hardening at about 350 GV of rigidity. These results confirm the early findings of the ATIC-2, CREAM and PAMELA experiments [23].

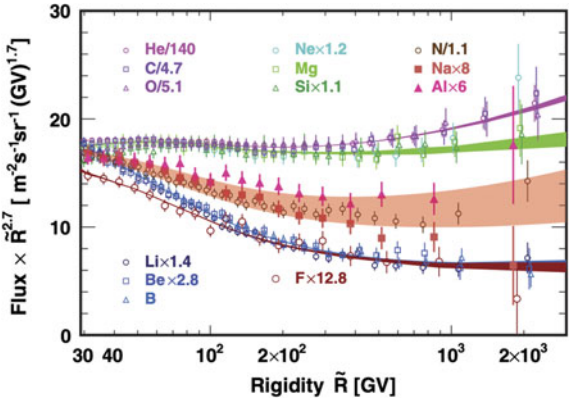

Fig. 4 The fluxes of cosmic nuclei measured by AMS as a function of rigidity from Z = 2 to Z = 14 above 30 GV. As seen, there are two classes of primary cosmic rays, He-C-O and Ne-Mg-Si, and two classes of secondary cosmic rays, Li-Be-B and F. Nitrogen (N), sodium (Na), and aluminum (Al), belong to a distinct group and are the combinations of primary and secondary cosmic rays. For clarity, data points above 400 GV are displaced horizontally. For display purposes only, fluxes were rescaled as indicated. The shaded tan band on N, Na, and Al is to guide the eye

Also, the proton-to-helium ratio at $R > 45$ GV if found to falls off steadily as p/He $R^{-0.077}$. Interpretations for these phenomena fall into three classes: diffusive shock acceleration mechanisms, propagation effects, or superposition of local and distant sources [24–27]. The origin of the CR spectral hardening (and its connection with the p/He ratio anomaly [28, 29]) is a open question that may be resolved with high-energy data on light CR nuclei. As of today, AMS has provide the measurement of:

- CR primary nuclei fluxes of Proton, Helium, Carbon, Oxygen , Neon, Magnesium and Silicon;
- CR secondary nuclei of antiproton, Litio, Beryllium, Boron and Fluorine;
- CR nuclei fluxes of Sodium, Nitrogen and Aluminium that are a combination of primary and secondary

Primary cosmic rays are thought to be mainly produced and accelerated in astrophysical sources. The precise knowledge of their spectra in the GV—TV rigidity range provides important information on the origin, acceleration, and propagation processes of cosmic rays in the Galaxy.

The AMS results shows that both primary and secondary nuclei fluxes deviate from a single power law above 200 GV, however, the rigidity dependence are distinctly different as shown in Fig. 4 that presents cosmic nuclei fluxes measured by AMS as a function of rigidity from Z = 2 to Z = 14. There are two classes of primary cosmic rays, He-C-O and Ne-Mg-Si, and two classes of secondary cosmic rays, Li-Be-B and F [30]. N, Na, and Al belong to a distinct group and are the combinations of primary and secondary cosmic rays. These results indicate there are two kinds of cosmic ray rigidity dependencies. These observations have generated new developments in

cosmic ray models [31, 32]. The theoretical models have their limitations, as none of them predicted the observed spectral behavior of the cosmic rays. The results on heavier primary cosmic rays Ne, Mg, and Si show that primary cosmic rays have at least two distinct classes of rigidity dependence. These unexpected results together with ongoing measurements of heavier elements in cosmic rays will enable us to determine how many classes of rigidity dependence exist in both primary and secondary cosmic rays and provide important information for the development of the theoretical models.

To examine the difference between the rigidity dependence of primary and secondary cosmic rays in detail, the ratios of the lithium, beryllium, and boron fluxes to the carbon and oxygen fluxes were computed. Precise knowledge of the charge and rigidity dependence of the secondary cosmic ray fluxes and the secondary to primary flux ratios is essential in the understanding of cosmic ray propagation. In particular B/C or the more direct B/O, have been traditionally used to study the propagation of cosmic rays in the Galaxy [32]. For the first time, AMS has provided the first measurement of He isotopes fluxes. This measurement can provide complementary information to the commonly used B/C or C/O ratio [33]. Helium nuclei are the second most abundant cosmic ray. They consist of two isotopes, ^{3}He and ^{4}He. ^{4}He are thought to be mainly produced and accelerated in astrophysical sources, while ^{3}He are overwhelmingly produced by the collisions of ^{4}He with the interstellar medium.

Helium interaction cross sections with the interstellar medium are significantly smaller than that of heavier nuclei. Therefore, helium travel larger distances, probing a larger galactic volume. Explicitly, the ratio ^{3}He/^{4}He probes the properties of diffusion at larger distances [34].

Protons and their anti-particle (anti-protons) can contribute to the indirect search of Dark Matter. The antiprotons constitute, as in the case of positrons, a rare component of CR: for each antiproton there are approximately 10^4 protons. Precision measurements of the cosmic ray antiproton flux are as important as measurements of cosmic ray positrons since both species are antiparticles that have to be created in high-energy processes rather then just being accelerated from the interstellar medium by a passing shock wave.

As discussed in Sect. 3.1, AMS has observed an excess in the positron flux. These data generated many interesting theoretical models including collisions of dark matter particles, astrophysical sources, and collisions of cosmic rays. Some of these models also include specific predictions on the antiproton flux.

As an example, Fig. 5 shows the AMS antiproton spectrum and the antiproton-to-proton flux ratio together with the four recent theoretical predictions for models with dark matter annihilation and cosmic ray collisions [35–37]. As shown the current uncertainties of the modeling need to be further improved before a definitive theoretical interpretation of the origin of cosmic ray antiprotons is possible. This shows the importance of comparing models with all the available AMS data, including also the data on electrons, positrons, protons, and nuclei. The accuracy of theoretical predictions can be improved with the latest AMS results on the fluxes of primary and secondary cosmic rays and their ratios to the models. The continuing AMS measure-

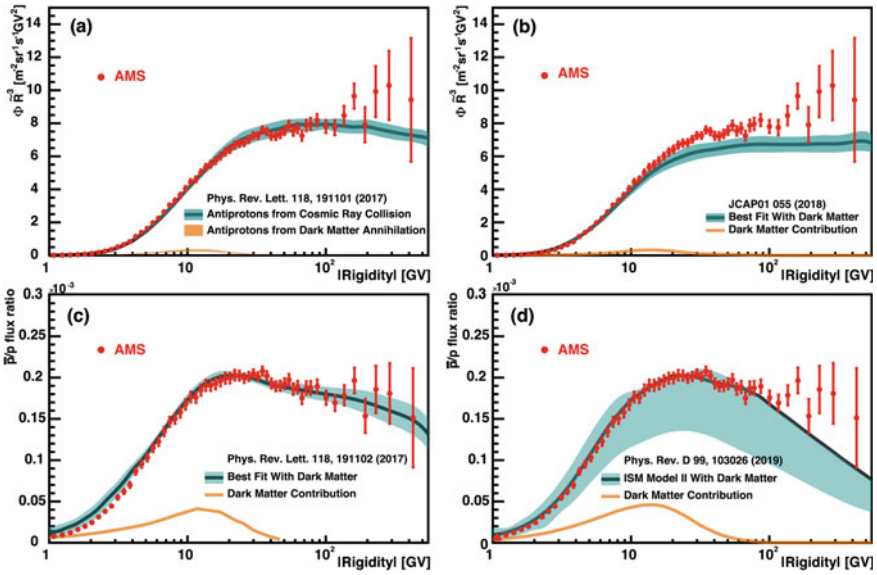

Fig. 5 **a**, **b** The AMS antiproton spectrum and **c**, **d** the antiproton-to-proton flux ratio (red data points) together with four recent theoretical models (black lines), their uncertainties (blue bands), and the contributions from dark matter annihilation (yellow lines) [35–37]

ments of the antiproton spectrum to the highest rigidity with improved accuracy, as well as its detailed time dependent variations, will also provide a crucial input to understanding of the origin of antiprotons in the cosmos.

4 Results on Solar Modulation of CR

In order to reach the Earth's atmosphere, the galactic CR have to pass through the heliosphere, interacting with the solar wind plasma and the solar magnetic field. We have the so-called *Solar modulation* (SM) of CR, that is a time, space, energy, and particle-dependent phenomenon that arises from basic transport processes of CR in the heliosphere [38]. As a result of SM, the measured flux below 30 GeV, is different from the galactic flux, i.e. not affected by the SM. The solar activity has a cycle which period is ∼11 years, during which it increases reaching a maximum and then decreases again. The intensity of cosmic ray radiation is correlated (or rather anticorrelated) with the activity of the sun. Another is the 22-year cycle of the Sun's magnetic field polarity, which reverses every 11 years during the maxima of the solar cycle and that gives a charge-sign dependent effect on CR fluxes.

The study of CR as a function of time and energy allows the study of solar modulation, i.e. of solar activity. Modelling of the heliospheric effects is a complex matter,

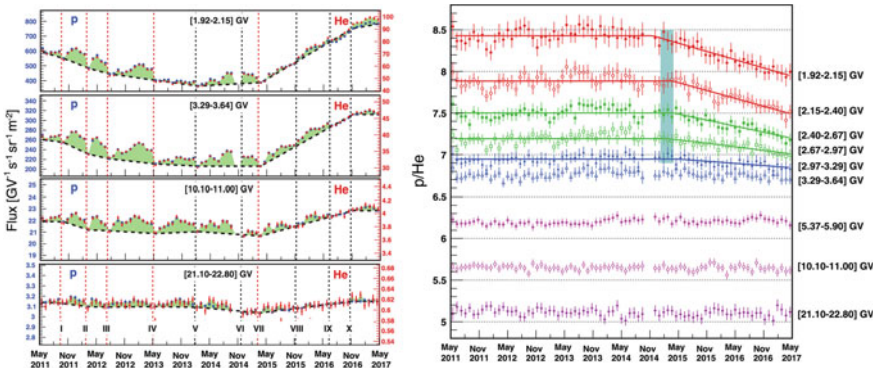

Fig. 6 On the left: The AMS proton (blue, left axis) and helium (red, right axis) fluxes as function of time for 4 rigidity bins. The error bars are the quadratic sum of the statistical and time dependent systematic errors. Detailed structures (green shading and dashed lines to guide the eye) are clearly present below 40 GV. The vertical dashed lines denote boundaries between these structures. On the right: the AMS proton/helium flux ratio as function of time for 9 characteristic rigidity bins. The errors are the quadratic sum of the statistical and time dependent systematic errors

as of today there are very few direct measurements of CR fluxes to correlate with continuous changing solar activities along the extended period of a solar cycle. The large acceptance and high precision of AMS allow us to perform accurate measurements of the fluxes as functions of time and energy with unprecedented precision over at least one complete solar cycle. As of today, AMS as already published the measurements of proton, Helium, electron and positron fluxes as a function of time, with a time granularity corresponding to one Bartels rotation (27 days) and is currently working on the measurements of fluxes in time of the other species of CR. This provides unique information to probe the dynamics of solar modulation, to allow the improvement of constraints for dark matter searches, to investigate the processes of galactic cosmic ray propagation, and to reduce the uncertainties in radiation dose predictions for deep space human exploration.

The time dependence of the proton and helium fluxes are shown in Fig. 6 (left) for 8 characteristic rigidity bins. As seen, both the proton and helium fluxes have fine time structures each with maxima and minima. The structures in the proton flux and the helium flux are nearly identical in both time and relative amplitude. In general, the amplitudes of the structures decrease progressively with rigidity. The precision of AMS enables us to observe these structures up to 40 GV. The amplitudes of the structures are reduced during the time period, which started one year after solar maximum (i.e. starting March 2015), when the proton and helium fluxes steadily increase. As shown in Fig. 6 (right), the p/He ration as a function on time is consistent with a constant above 3.29 GV. Below 3.29 GV, the observed /He flux ratio is steadily decreasing with time after the solar maxima. This behaviour it has been observed for the first time and shows a new and important feature regarding the propagation of

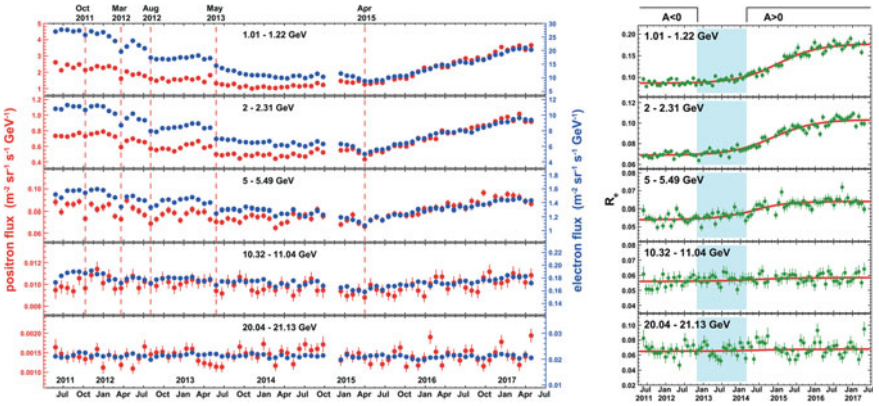

Fig. 7 Left: Fluxes of primary cosmic-ray positrons (red, left axis) and electrons (blue, right axis) as functions of time, for five characteristic bins. The error bars are the statistical uncertainties. Prominent and distinct time structures visible in both the positron spectrum and the electron spectrum and at different energies are marked by dashed vertical lines. Right: The ratio R_e of the positron flux to the electron flux as a function of time. The error bars are statistical. The polarity of the heliospheric magnetic field is denoted by $A < 0$ and $A > 0$. The period without well-defined polarity is marked by the shaded area

lower energy cosmic rays in the heliosphere. The precision of the AMS data provides information for the development of refined solar modulation models [39, 40].

Figure 7 (left) shows the AMS measurements of the time and energy dependence electron and positron fluxes as a function of time for five characteristic energy bins. Both electron and positron fluxes show short-time structures, on the time scale of months. The data show a clear evolution of the fluxes with time at low energies that gradually diminishes towards high energies [**?**]. Both fluxes exhibit profound short- and long-term variations. The short-term variations occur simultaneously in both fluxes with approximately the same relative amplitude. Viceversa, the long-term time structure appear with different relative amplitude for electrons and positrons. This can be clearly seen from Fig. 7 (right) that shows the electron/positron ratio as a function of time in five characteristic energy bins of the fluxes. The short-term variations in the ratio largely cancel, and a clear overall long-term trend appears. At low energies, is flat at first, then smoothly increases after the time of the solar magnetic field reversal, to reach a plateau at a higher amplitude.

Since electrons and positrons differ only in charge sign, their simultaneous measurement offers a unique way to study charge-sign dependent solar modulation effects. For the first time, the charge-sign dependent modulation during solar maximum has been investigated in detail by leptons alone. These data allow comprehensive studies of the energy and charge-sign dependence of short-term effects on the time scale of months, related to solar activity [41, 42], and long-term effects on the time scale of years, related to the 22-year cycle of the solar magnetic field polarity.

5 Conclusions

The AMS experiment has collected more then 190 billion cosmic rays and it will continue to collect data for the lifetime of the International Space Station (2028 and beyond). The current AMS results on CR bring to a puzzled scenario. Electrons and positrons of high energies originate from different sources. In particular, the high energy positrons source, seems to be compatible both with Dark Matter and Pulsars hypothesis. In the future, collecting more statistics, AMS will be able to shed a light on the nature of the source.

The AMS results on CR nuclei, show that primary and secondary nuclei have at least two distinct classes of rigidity dependence but that the rigidity dependence of the two secondary classes is distinctly different from the rigidity dependence of the two primary classes. These are new and unexpected properties of CR. These unexpected results together with ongoing measurements of heavier elements in cosmic rays will enable us to determine how many classes of rigidity dependence exist in both primary and secondary cosmic rays and provide important input to the development of the theoretical models.

For the first time, AMS allows the investigation of charge-sign dependent modulation effect during solar maximum by leptons alone with the measurement of time evolution of electron and positron fluxes. The AMS proton and helium fluxes, show a different behaviour in time below 3.29 GV. This behaviour it has been observed for the first time and shows a new and important feature regarding the propagation of lower energy cosmic rays in the heliosphere. The understanding of solar modulation effects on CR, is fundamental not only to have a correct interpretation of the measured CR fluxes but also to addresses a prerequisite for modeling space weather effects, which is an increasing concern for space missions and air travelers. The study of these effects has been limited for long time by the scarcity of long-term CR data on different species, and by the poor knowledge of the local interstellar spectra. A continuous stream of time-resolved and multichannel CR data is being provided by the AMS.

AMS will continue collecting data through the life of the ISS exploring the physics of complex anti-matter (anti-He, anti-C, etc.), the physics of dark matter (anti-deuterons, anti-protons, and positrons), the physics of cosmic-ray nuclei across the periodic table, and study solar physics over the entire solar cycle.

References

1. P. Blasi et al., Astron. Astrophys. Rev. **21**, 70 (2013)
2. P. Chardonnet et al., Phys. Lett. B **409**, 313–320 (1997)
3. A.G. Mayorov et al., Bull. Russ. Acad. Sci.: Phys. **75**(3), 2011
4. M. Aguilar et al., Phys. Rev. Lett. **114**, 171103 (2015)
5. O. Adriani et al., Phys. Rev. Lett. **111**, 081102 (2013)
6. M. Ackermann et al., Phys. Rev. Lett. **108**, 011103 (2012)
7. C. Grimani et al., Astron. Astrophys. **392**, 287 (2002)

8. M. Boezio et al., Adv. Space Res. **27**, 669 (2001)
9. M. Aguilar et al., Phys. Lett. B **646**, 145 (2007)
10. S.W. Barwick et al., Astrophys. J. **498**, 779 (1998)
11. M.A. DuVernois et al., Astrophys. J. **559**, 296 (2001)
12. M. Aguilar et al., Phys. Rev. Lett. **122**, 041102 (2019)
13. C.H. Chen et al., Cosmol. Astropart. Phys. **03**, 041 (2017)
14. Y. Bai et al., Phys. Rev. D **97**, 115012 (2018)
15. N. Tomassetti et al., Astrophys. J. Lett. **803**, L15 (2015)
16. D. Hooper et al., Phys. Rev. D **96**, 103013 (2017)
17. S. Profumo et al., Phys. Rev. D **97**, 123008 (2018)
18. D. Hooper et al., Cosmol. Astropart. Phys. **01**, 025 (2009)
19. T. Linden et al., Astrophys. J. **772**, 18 (2013)
20. M. Graziani et al., Proc. Sci. **367** (2019)
21. M. Graziani et al., J. Phys. Conf. Ser. **1468**, 1 (2020)
22. M. Aguilar et al., Phys. Rev. Lett. **115**, 211101 (2015)
23. O. Adriani et al., Science **69**, 171103 (2011)
24. A.E. Vladimirov et al., Astrophys. J. **752**, 68 (2012)
25. N. Tomassetti et al., Astrophys. J. **803**, L15 (2015)
26. R. Aloisio et al., Astron. Astrophys. **583**, A95 (2015)
27. N. Tomassetti et al., Phys. Rev. D **92**, 081301(R) (2015)
28. N. Tomassetti, Astrophys. J. **815**, L1(R) (2015)
29. Y. Ohira et al. (2015). arXiv:1506.01196
30. M. Aguilar et al., Phys. Rev. Lett. **126**, 081102 (2021)
31. M.J. Boschini et al., Astrophys. J. **840**, 115 (2017)
32. C. Evoli et al., Phys. Rev. D **99**, 103023 (2019)
33. I.A. Grenier et al., Annu. Rev. Astron. Astrophys. **53**, 199 (2015)
34. G. Jóhannesson et al., Astroph. J. **824**, 16 (2016)
35. A. Cuoco et al., Phys. Rev. Lett. **118**, 191102 (2017)
36. I. Cholis et al., Phys. Rev. D **99**, 103026 (2019)
37. M.Y. Cui et al Phys. Rev. Lett. **118**, 191101 (2017)
38. M.S. Potgieter, Living Rev. Sol. Phys. **10**, 3 (2013)
39. L.F. Burlaga et al., Astrophys. J. **407**, 347 (1993)
40. Jr.G. Newkirk et al., Res. Space Phys. **86**, 5387 (1981)
41. H.V. Cane, Space Sci. Rev. **93**, 55 (2000)
42. M.S. Potgieter et al., Astrophys. J. **403**, 760 (1993)

The Right Key

Four Spacewalks to Repair the Alpha Magnetic Spectrometer on the International Space Station

Claudio Bortolin and Paola Catapano

At the beginning of 2014, right after the end of my first mission, a NASA engineer asked for my opinion about the feasibility of carrying out a complex mechanical repair on the AMS-02 experiment. She showed me some slides where the problem was described, asking for my opinion. I replied that, since it was an engineering intervention, we had to adopt an engineering approach, so with the right tooling we would be able to do it. At that time, I would never have expected to be the one who would carry out the activity.

This is how Italian European Space Agency (ESA)[1] astronaut Luca Parmitano recalls the very first time he was asked to comment on the repair of the CO_2 cooling system of the AMS-02 cosmic detector, which is installed on the outside of the International Space Station (Fig. 1) [1]. It was February 2014, when the whole operation was still just under evaluation. And who would disagree with Colonel Parmitano when he said: "Repairing any part of a particle detector is a complex and delicate operation even on Earth, now let's imagine having to do it in Space, outside the ISS and wearing pressurized gloves."

Every physicist and engineer who deals with particle detectors and their infrastructure is used to dealing with unprecedented issues during the operation of this sophisticated custom-made equipment. For instance, the thermo-mechanical stability is a key parameter for the trackers—the ultralight innermost subsystems of a particle physics detector. We know that even just one-degree difference in temperature between the opposite sides of a silicon sensor can affect the micrometric precision of the particles'

[1] European Space Agency https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/Astronauts/Luca_Parmitano.

C. Bortolin (✉)
Experimental Physics Department, ATLAS Detector Operation/ Project Office (EP/ADO/PO), CERN, Route de Meyrin, 1211, Geneva 23, Geneva, Switzerland
e-mail: Claudio.Bortolin@cern.ch

P. Catapano
Communications & Outreach Group, International Relations IR/ECCO/ECP), CERN, Route de Meyrin, 1211, Geneva 23, Geneva, Switzerland
e-mail: Paola.Catapano@cern.ch

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_13

Fig. 1 The International Space Station (Photo credit NASA)

path reconstruction, thus disrupting the description of the physical phenomena that caused the interaction. The work of thousands of scientists would be jeopardized by that tiny detail!

Cooling engineers at CERN would unquestionably agree with Luca Parmitano: repairing any detail of a cooling structure, which is integrated in a fragile detector structure, isn't at all an ordinary operation, even on earth. As scientists operating some of the most extraordinary scientific tools humankind has conceived and built, we are often required to think out of the box, to find the solutions that are needed to repair or upgrade technical equipment. This is what happened to CERN engineer Claudio Bortolin in 2011, when he was a young collaborator in the ALICE experiment, one of the four giant detectors installed at the LHC,[2] some eighty meters underground. A stainless-steel twisted cable—the same used in bicycle brakes—was the key "tool" he used to develop a flexible drilling system that he and his team used to drill ten clogged filters of the cooling circuits of the silicon detector, which was situated four meters away from the last access point to the underground area. The elements to be drilled were at the end of a four-millimetres stainless-steel pipe with a couple of corners on the way [2]. Any alternative solution would have involved dismantling the ten-thousand ton equipment to reach that tiny filtering element—certainly not a simple operation!

[2] Large Hadron Collider.

Those clogged filters weren't for Claudio just a personal success but also the reason why, on a Saturday afternoon of May 2015, he received a phone call, informing him of the cooling issue on AMS-02.

1 Back to the Beginning

On that sunny afternoon Claudio was, as usual on most Saturdays, guiding the tour of a group of Italian high-school students visiting CERN. By coincidence, they were at the Payload Operation Control Center, or POCC, the "control room" of AMS-02, one of Claudio's favourite visiting sites on offer to CERN visitors. Claudio was fascinated by space research since his early youth, and quite naturally, this had become one of the favourite topics of his guided tours. He was particularly good at making visitors feel part of the human adventure called "Science", by adding anecdotes and personal stories to his presentations. On that Saturday afternoon, he was explaining some details of the CO_2 cooling system of the silicon detector of AMS-02, a very similar system to the ones he was operating and maintaining in the experimental caverns at CERN, when his mobile phone rang. The display read "Bart Verlaat", a colleague of his. The unexpected call on a Saturday worried him because they were both involved in many activities related to CO_2 cooling systems installed in the LHC caverns, and the call sounded like a red alert. Before picking up, he told the group of students: "*Wow, what a coincidence, the Dutch engineer who made the first drawing of AMS-02 is just calling me!*" Indeed, Bart is the most experienced expert in CO_2 detector cooling systems at CERN. "*Professor Samuel Ting*[3] *called me last night and asked me to get on the first plane to Geneva because we have a problem with the AMS CO_2 cooling pumps*" – he informed Claudio. "*I need your help with the fiberscope you used a few years ago to investigate the problem of the blocked filter in the ALICE experiment. We have a replica of the AMS pump at CERN, I plan to check it by looking inside.* "Claudio looked at the young visitors: "*Do you remember the famous sentence*: ***Houston we have a problem?*** *Well, this is one of those moments:* ***Houston we have a problem on AMS***."

Was the problem at all expected? In 2015, NASA had decided to extend the International Space Station program to 2024; in 2018 the program was extended to 2030. AMS-02 was initially due to run until 2020, but the program extension was felt by the AMS-02 Collaboration as a great opportunity to run the experiment a few years more. However, after a first analysis of the pressure trends on the AMS cooling system and the investigations on the pump replica, at the end of May 2015 it became clear that, in order to operate AMS for as long as the ISS was running, it would have been necessary to replace part of the cooling system, in particular the four pumps that had shown the first signs of failure.

[3] Prof. Samuel Ting is the 1976 Nobel Laureate in Physics and the principal investigator of the AMS-02 collaboration.

As Nobel laureate and AMS principal investigator Sam Ting said during an interview, "*The first time something was wrong it was February 2014, when a pump stopped. We had three functioning, but I called NASA and said, if a pump stops, you can't assume the others will not. We needed to start thinking of a replacement. There was tremendous resistance, mostly technical, because the pumps are connected with high pressure lines and we didn't know how to start, we didn't know what was wrong with the pump* [3]."

Soon after May 2015, a team of engineers and physicists from CERN and NASA, began to study the problem and set up the repair procedure for an experiment that wasn't designed to be repaired. Actually, it wasn't just a matter of identifying and testing one single procedure, but rather a set of very sophisticated and unprecedented procedures, such as reaching the working site of the experiment on the ISS, or accessing the cooling system, removing the thermal insulation called MLI (Multi-Layer insulation), cut the pipes, connect a new system and the relevant power supply and sensors reading cables, and finally run the pressure test and put the system back into operation. Obviously, all this work had to be divided into more than one Extra Vehicular Activities (EVAs). Eventually, four EVAs were established, each lasting about ten hours, with the astronauts wearing a pressurized spacesuit while traveling at 28 thousand km/h around the planet at more than 400 km of altitude. Nothing similar had ever been attempted before! Even for CERN engineers, with the experience accumulated in experimental areas over several decades, this looked like a real challenge.

> *This repair was very difficult*", continued Sam Ting during the same interview. "*It can be compared to a heart transplant. It took four EVA's, originally it was planned they would be five or six. These were the most difficult EVA's NASA ever took, and it took us five or six years to prepare them. First thing, we ordered an identical pump and by checking its functioning we concluded that replacing the pump was not enough, we would need to replace the whole central part of the cooling system. It's like taking a whole heart out and replace it entirely and reconnect it to the body. NASA has been extremely supportive. It was not a question of money but ability. How to do this? When we started the system, we never thought one day we would replace it entirely. We have four pumps, so there's a 400% redundancy. Last year, we were only at 50% efficiency, so it was a good time to replace it.*

Imagining, carefully designing and finally testing all these procedures, required a precise and meticulous preparation that lasted about five years and involved many engineers, technicians and some astronauts, who practiced the whole operation over and over again, in various simulated environmental conditions, recreated on an AMS-02 mock-up installed on the ISS truss structure and at the NASA Neutral Buoyancy Lab in Houston. In mid-November 2019, a rocket launched from Virginia was the last of four launches required to bring all the materials needed for the AMS EVAs to the ISS, including a white box with the brand-new pump system, without which AMS-02 would have been left with no chance to continue its data taking activity and potentially make new discoveries. Together with the new cooling system, that rocket transported also twenty-nine special tools, custom-designed and developed by the expert team specifically for the AMS EVA programme. They included a commercial tool, a 7/8″ hexagonal key, technically called wrench, that Luca wanted to be photographed with

before getting back to the airlock of the Space Station. That wrench made history. Parmitano published its photo on his Twitter account a few hours after the EVA with the following text: "*My favourite picture of last Saturday's EVA, so symbolic: that wrench saved the day - and AMS!*" But the story of that 7/8″ hexagonal wrench had started long before, in Aachen (Germany), in a small workshop in the basement of the apartment where Ken Bollweg, the engineer at NASA/JSC AMS-02 Project Office team, was living during the construction of the new cooling system. Talking with Ken, it becomes immediately clear that mechanical engineering is part of his DNA: "*My father had a professional activity in the plumbing, electricity and heating sector and I was 11 when I started handling mechanical equipment. At 12, I built my first pump, while at the age of 14 I put together my first car engine. AMS-02, in the end, is just about plumbing, but it has to be carried out in space* [4]."

2 The Key Tooling

The central part of AMS-02 consists of a permanent magnet whose role is to bend the tracks of cosmic particles crossing its magnetic field with an intensity of 1500 Gauss (more or less 3000 times higher than the Earth's magnetic field). The particles' charge can then be inferred by observing the direction of the bended tracks, while the radius is proportional to the momentum of the particle. On AMS-02, the tracking capacity is provided by nine silicon detector planes. In total, the path of the particles from the entry point to the exit point is about three meters, the sum of the area of each silicon plane is about seven square meters, and the precision of the particle track reconstruction is about ten microns (10 millionths of a meter). Without its silicon tracker, AMS-02 couldn't not continue to take data and do the important science it is doing. The CO_2 cooling system is an essential element of the tracking detector operation, as it guarantees the required thermal stability for the silicon sensors to work. The one-degree tolerance must be guaranteed in spite of the harsh outdoor environmental conditions of the ISS, where temperature can vary from −160 °C at night to +120 °C during the day, at each orbit, which lasts about 90 min. So, when operators observed the first signs of aging of the cooling pumps, four in total, alternating one in run and three in stand-by, they had little doubts: in order to extend the lifetime of AMS-02, and even if the experiment was not designed and built to be repaired once installed on the ISS, there was no alternative but to replace a large part of the cooling system.

The preparation of the EVAs required the development of several tools, which were needed to remove the MLI, get access to the stainless-steel pipes, cut them, install new connectors with front and back ferrules to engage the pipe. All this, obviously, while making sure the whole system was leak-tight. And with the astronauts having to carry out all these operations from inside a pressurized spacesuit and wearing pressurized gloves. To get a better idea, we should imagine ourselves trying to repair a wristwatch operating on the small cogwheels while wearing ski gloves [2].

At the RWTH[4] University of Aachen in Germany, experts were working on developing and testing the new cooling system (Upgraded Tracker Thermal Pump System—UTTPS); in the meantime, another team in Houston was preparing the tools that the astronauts would have to use during the EVAs. Ken Bollweg had participated in the AMS adventure as senior engineer for NASA since the 90s, when AMS-01, a precursor experiment, flew onboard Space Shuttle Discovery for 10 days in June 1998 (STS-91 mission). During this test mission, AMS-01 collected enough cosmic-rays data to demonstrate the feasibility of building, transporting and running an experiment for high-energy physics in space. Ken was part of the so-called APO, the AMS Project Office, the group responsible for payload integration, safety, and oversight of AMS operations on the Space Shuttle and the International Space Station. To explain the complexity of the EVAs, Ken mentioned the long procedures that astronauts have to follow many times during their training, before they can carry them out at best in space: "*In every EVA there are procedures to follow, pages and pages of instructions that increase a lot if they include the steps to solve multiple issues. A fundamental aspect was added to the EVAs of AMS-02: the experiment was not designed and built to be repaired in space. Despite this fundamental aspect, there were only two cases during four EVAs where astronauts had to apply a corrective procedure, something absolutely amazing, unprecedent.*" In a complex operation such as the repair of AMS-02, the preparation was absolutely the most complex part, the tools developed, and the many hours dedicated to training, to use and improve them, were key factors: "*It required an enormous effort of preparation and interaction to verify and improve the procedures; we used a new facility called ARGOS, Active Response Gravity Offload System, a system based on a cable supporting the astronaut that interacts with a mobile platform to simulate reduced gravity environment. It allowed us to closely observe the astronauts while they were exercising and we could realize what aspects we could improve, both in terms of procedure and on the tool itself.*"

While some NASA engineers focused on the possible use of new connectors to facilitate the work of the astronauts, Ken decided to investigate an alternative route. His idea was to continue using the same commercial connectors already installed and performing well on the AMS-02 cooling circuits (the same connectors that are widely used at CERN). He would then integrate them with the ferrules in a single mechanical tool to make installation easier and at the same time enable the astronauts to check the pressure tightness without any additional tools. Since he was also following the construction and testing of the UTTPS, Ken decided to build a small mechanical workshop in the basement of the building where he had rented an apartment during his stay in Aachen.

In a perfect NASA style, Ken Bollweg began to develop this tool which then, thanks to the collaboration of institutes and private companies such as NASA-APO (Houston—US), Jacobs (Dallas-US), MIT (Boston-US), INFN-SERMS (Perugia—Italy), RWTH (Aachen—Germany) and (HAKU GmbH Aldsorf-Germany), became what has been called AAF—the AMS Advanced Fitting. The Fitting is a cylindrical

[4] Rheinisch-Westfälische Technische Hochschule Aachen is a German public research University located in Aachen, North Rhine-Westphalia, Germany.

shape mechanism, in which two pipes can be connected just by using one hand to hold and an hexagonal wrench to close, at the same time. Inside the Fitting, a pre-assembled connector, made of male, female and ferrules, was ready to be swaged. The way in which these elements are tightened is especially important for their functionality. If they are swaged too little or too much, there is the risk of a leak. The optimal rotation angle to swage them is obtained with one complete turn of 360 degrees. This is the reason why some letters and numbers were printed in the central part of the AAF body: to allow astronauts to check the rotation angle while swaging the external nut by using the famous 7/8 "hex wrench. They would take note of the letter and the number and then use the wrench to swage one complete turn, until the same letter and number were aligned again. Another crucial purpose of the AAF was the leak check, to be performed after tightening the connectors. A small piston integrated in the cylinder would be triggered in case of any pressurization of the volume due to a CO_2 leak. The system is called the Visual Leak Indicator (VLI). Luca Parmitano and Andrew Morgan, the astronauts engaged in the EVAs for this mission, were trained to deal with any event in which the piston would move and a red band would appear. Such training took place in Houston, at a mock-up of AMS-02, where the tools prepared for the EVAs programme were flown to start the testing phase of the entire operation.

3 Planning and Preparing

In space, any unexpected situation is a source of risks which must be minimized. The success of the mission is strongly linked to the training on the ground before the EVAs. To leave nothing to chance, several astronauts got involved during this phase of the mission preparation. Their invaluable experience was a key contribution to the improvement of the tools, the design of the movements to carry out, the exact sequence of the procedures and eventually the operation of the equipment. Veteran astronaut Chris Cassidy (retired on 28th May 2021) tested the procedure twenty-nine times! The exercise took place partially inside the flotation pool at NASA's Neutral Buoyancy Laboratory—a diving tank of 62 m length, 31 m wide and more than 12 m deep. The role as lead Spacewalker (EV1) is usually assigned to astronauts who have already gained extravehicular experience, so it was decided to assign it to Luca Parmitano; for Andrew (called usually Drew) Morgan, a physician of the US Special Forces Group recruited in 2013, this was the first mission in space. Luca completed fourteen runs of training, twelve of which as EV1 and two as EV2 while Drew completed all his eight exercises as EV2. Other astronauts were trained and repeated the procedure over and over again, to refine it in detail and to find any small problem or identify any risks: Michael Hopkins did twelve runs, Jeremy Hansen eleven, Kathleen Rubins four, Anne McClain two. A further nineteen runs were carried out by another group of astronauts, for a total of ninety-five runs. Each time, the full exercise was tested, improved and optimized, the tools were tested and the backup scenarios implemented. The date of launch was approaching in the meantime. On July 20th, 2019 space mission "**Beyond**" was about to begin.

Fig. 2 Andrew Morgan and Luca Parmitano check spacesuits and tools in the Quest airlock (Photo credits NASA)

4 Lift-Off and the "Opening Act"

July 20th, 2019, is an important day for all of humanity. The fiftieth anniversary of the first human on the Moon is celebrated in every corner of the planet as a lasting symbol of Space conquest. The Apollo 11 mission is part of the imagination of all the children of the Earth; every child born after 1969 has at least once dreamt to become like Neil Armstrong, when he was photographed by Buzz Aldrin on the surface of the Moon, while Michael Collins was orbiting around it, inside the Command and Service Module (CSM). Seven years later, two of those children were born and thanks to their commitment, dedication and a little luck, their dreams became true: Luca Parmitano and Andrew Morgan. On that day, they were on board the Soyuz MS-13 spacecraft with Russian commander Aleksander Skvortsov, when it took off from Baykonur (Kazakhstan) at 16:28 (UTC) to dock the International Space Station about six hours later, at 22:47 (UTC). When the last payload with spares and tools for AMS-02 docked on the ISS, came the day of the first extravehicular activity (Fig. 2).

During an interview to PassioneScienza.com,[5] Luca, wearing a polo shirt with the AMS-02 logo, shared a detail that helps to understand the team spirit of all those involved with the AMS-02 repair operations.

[5] PassioneScienza.com is Claudio Bortolin's webplatform, where he streams innovative science shows targeted specifically to school students.

We decided to give nicknames to each of the EVAs [5]. *For the first one, we decided to call it "The Opening Act", which in English can also be called "Overture", a pun that well described the open-heart operation that we were about to begin. We literally had to create an access to the detector, by removing the protective layers, thus opening a passage and getting to the area where subsequent operations would be carried out in the following EVAs.*

The difficulty of the AMS-02 operation was compared to the repair of the Hubble Space Telescope in the early 1990s. The space station, including its solar panels, is about the size of a football field, 100 m long and 80 m wide. The pressurized and inhabited part of the station is located in the centre, from where the main truss structures develop and branch off along the side of the station. Along these trusses, various payload connecting points are located, supporting structures where experiments in vacuum are carried out, such as the ExPRESS Logistics Carrier (ELC), or storage areas of components such as the External Storage Platform (ESP). AMS-02 is mounted on one of these points, approximately twenty-two meters from the pressurized zone. This was the distance that Drew had to travel repeatedly, moving by hand along the truss structure between the Airlock and the working area, and carrying the necessary equipment back and forth several times. In the meantime, Luca was hooked to the CanadArm2, manoeuvred from inside the Space Station by fellow astronauts Cristina Koch and Jessica Meir, who moved Luca directly to the AMS-02 detector [6].

The first task to be carried out once we reached the working site, was to unscrew 27 bolts, to be able to remove a shield panel, a protective metal element behind which we would have had access to the pipes of the cooling system. The bolts used for AMS-02 obviously had not been chosen with a possible repair in orbit in mind. The ones we use during the EVAs are especially made so that screws and washers cannot be dropped, which would be dangerous for the ISS and for the astronauts. The twenty-seven bolts to be removed from AMS-02 were all of different sizes and, of course, came with washers and nuts. We had to use specially designed tools, installed on what we called the Power Grip Tool or PGT, an electric unscrewing system, in order to capture these small objects which, if lost, might become potentially dangerous. There wasn't enough room to use the PGT everywhere, so I had to remove some bolts by hand. Furthermore, before extracting the metal panel, to be able to handle it safely with pressurized gloves, we mounted a handling mechanism. Once disassembled, I passed it on to Drew who threw it down in a dispersion cone, a specific area identified by the Earth team, where the panel gradually lost altitude and eventually self-destroyed once in contact with the Earth's atmosphere.

An operation such as the one described by Luca would have been all in all trivial on Earth and, even in case of loss of a washer, one would not have worried at all. But on the orbiting station it is anything but simple. It is interesting to notice that the automatic screwdriver used by Luca was the successor of a similar tool used to repair the Hubble Space Telescope several years earlier.

Once the cavity inside which the pipes were located was opened, we began to remove the MultiLayer Insulation or MLI. These are layers of thermal insulation (necessary to protect the pipes from external temperature variations) that we had to remove to have access to the pipes, that we would then cut.

In total, the astronaut had to cut eight pipes to isolate the old system and connect the new one; six of these could be reached from the newly opened panel, while the

other two were located on the lower part of the experiment, even more complicated to reach. All tubes had an outside diameter of 4 mm with an inside diameter of 3 mm. Once the MLI was removed, Luca and Drew finally had access to the Vertical Support Beam (VSB), a U-shaped carbon fiber structural beam, where the tubes were housed, protected by a carbon fiber cover.

The idea of using the VSB as an access point to the pipes came from Italian engineer Corrado Gargiulo, a CERN staff member, who during the construction of AMS-02 was responsible for the integration of the experiment [7]. If there is a person who knows where each bolt is placed and how to get there, that's him: "*It is thanks to the procedures required by NASA, that we have been able to put together a reasonable procedure to connect a new cooling system to an existing experiment. Also, at CERN we usually document a lot of what is in the experiments, the various equipment, infrastructures and instrumentations, but the level of detail required by NASA is a different thing. The AMS-02 experiment is made up of thousands and thousands of elements, the effort required was enormous: from the nomenclature of the functions, the preparation of documentation regarding the position and size of the elements, together with their photos. However, this effort paid off years later! It is thanks to that information that I proposed to access the pipes from that spot, the VSB was well suited for cutting and connecting the new system. Other access points were considered during the study of the intervention, but at the end I have suggested that the VSB would be the best option based on the documentation we prepared ten years earlier.*"

The proposal was well accepted by NASA, given that any other location for accessing the pipes had been discarded. Corrado points out that that decision still involved a huge amount of work before validation: "*It is starting from my proposal that 29 tools were designed, developed and validated through multiple extensive trainings with the astronauts, an incredibly complex and successful job.*" However, access to the VSB was not without its problems, including safety, since manipulating carbon fiber surfaces with pressurized gloves involves quite some risks. The tension was palpable both in the control room in Houston and at CERN's POCC.

Luca Parmitano recalls those moments, trying to convey the reasons behind the tension. "*That cover wasn't just bolted on, we literally had to break it off, in order to take it off. It was thin enough to pull it off, but imagine grabbing a potentially sharp object with pressurized gloves! The operators at mission control were very concerned about this task, but it went well and after removing the cover, Drew used the same dispersion cone to launch it and let it degrade in the atmosphere, until it destroyed itself like the previous panel. The last task of this first EVA was to arrange the new power and data transfer cables close to the connectors in the old cooling system. With this, we concluded the first extravehicular activity and returned to the Airlock.*"

5 The Nail-Biter

A week after the Opening Act, Luca and Drew were back in the Airlock to start the next stage, the second extra-vehicular activity, which took place on November 22nd, 2019.

> *The nickname chosen for this second EVA was "The nail-biter", a way to represent the pressure the team was under, since it was time to cut the pipes of the old cooling system. This was a complex operation, since we first had to identify which pipes to cut: they were all the same diameter, and looked identical. Moreover, we had to cut them in the right sequence and at the right place. There was clearly a lot of tension in the team that day, because any mistake would have effectively terminated the operation and we would have had to start all over again, including the tools preparation on Earth.*

AMS-02 that day was disconnected from the system that had allowed it to work until a few days earlier, an error would mean a long interruption or even the end of the data taking, most likely for good.

Inside the VSB there were ten pipes of 4 mm in diameter, but only six of them were to be cut; the remaining four were arranged in a continuous pattern, from both ends, and the risk of cutting the same pipe in two different points was therefore very high. "*The pipe cutter used to perform the cut was a commercial tool re-adapted for space operations; once cut, the pipes became like real darts, and there was a serious risk of cutting the pressurized space suits, thus endangering the astronauts themselves, in this operation. Before getting to the more stressful moment, we had to prepare the data signal connectors between the new and the old system. While the new connectors were suitable for their use during extra-vehicular activities, those of the old system could damage the spacesuit if touched and for this reason I had to use a special tool to hold the old connector with a glove, while I connected the new connector with the other hand. Carrying out these tasks using pressurized gloves isn't easy at all, it requires considerable force to be able to exert any pressure, it is not difficult to imagine how much effort is required just to hold small objects like these in the hands. Fortunately, intense training runs on the ground enabled me to acquire the required manual skills."*

After installing a mechanical interface with the PGT, that would later enable to hook the box containing the new cooling system, came the most critical moment: the cutting of the pipes. *"Identifying the pipes before cutting them was a painstaking job! Since the pipes are identical, we had to be absolutely sure to use the pipe cutter on the correct ones. Once cut, I folded the ends of the tubes outwards and plugged them with special mushroom-shaped caps with the corresponding number marked on. These would allow me to avoid accidental collisions with sharp objects and then to easily identify them for their connection to the new system. The work area around AMS-02 was full of very fragile structures and sensors, so another complication was to avoid any accidental collision with this equipment which, if damaged, could have meant the end of the experiment. So, sometimes I was working in a very constrained position, movements had to be very precise and controlled. After cutting the pipes and extracting the VSB, I had to close the area with the MLI to protect the internal*

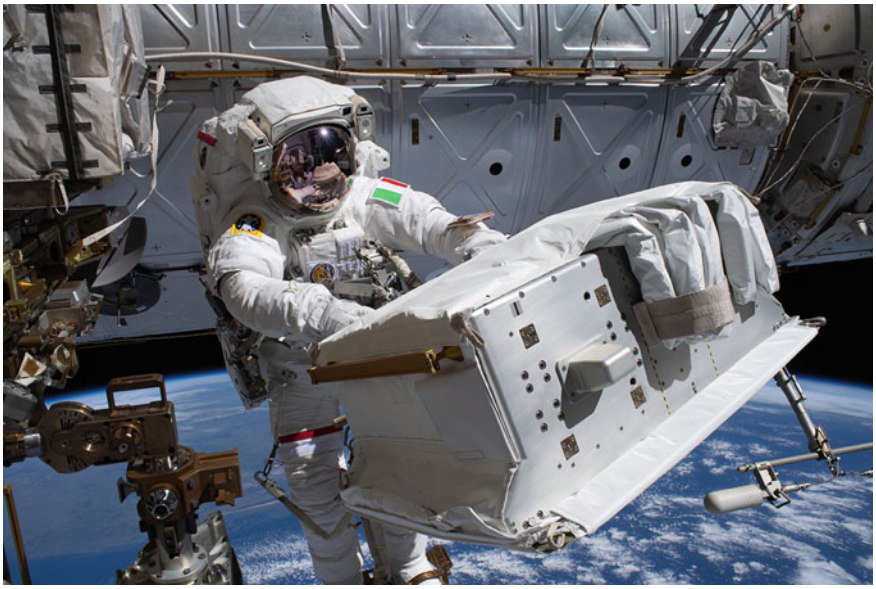

Fig. 3 Luca Parmitano outside the ISS with the box containing the new pumps (Photo credit NASA)

part from the external environment. With this operation the second extravehicular activity was completed" (Fig. 3).

6 The Money Run

It was December 2 when what Luca called the moment of truth came, that of the third spacewalk.

> *Here we are, 'the money run' name was decided by the team precisely to convey the meaning of a bet, a "failure or success" operation. I was taken to the work area with the robotic arm, while holding the UTTPS, the box with the new cooling system, which had been handed over to me by Drew shortly before. The first thing Drew and I did was to hook the UTTPS on the interface we had mounted during the second EVA; only at that point we would have started the much more complex phase, that is to connect the cut pipes of the old system with new ones.*

There are several tools on the market for cutting steel pipes, the most precise and accurate ones are pipe cutters designed to rotate around the pipe, so as to cut it little by little, keeping a nice round shape. In the case of AMS-02, such a tool could not be used, since there was no room available behind the pipes. So, Luca used a sort of shear, especially developed and adapted by NASA engineers. However, the effect of using a tube shear is that the tube gets pinched in the cutting area.

The first operation was to cut the pipes again using a rotational pipe cutter; if I could have rotated it around the pipe, it would have taken 22 turns to complete the operation and get what in technical terms is called a "clean cut", giving the tube its original shape with regular cutting edges. In the specific conditions of the AMS pipes, I could apply at most three quarters of a turn, so I had to repeat this manual operation many more times to get a clean cut comparable to the 22 full turns theoretically required. It was a rather long and tiring operation, flexing your fingers to use the tools with pressurized gloves is like trying to squeeze a tennis ball. Each tube required about two minutes of work; doing it on eight tubes was a rather tiring part.

When two pipes have to be connected, another important factor is their alignment; in fact one of the risks is that leaks will occur once the fittings are mounted. Some tools that help straighten the tubes are available and, exactly like on Earth, this operation was required also on the orbiting station.

Another very important phase began immediately after: we had to make sure that each pipe was straight enough to be able to insert it into the AMS Advanced Fitting (AAF). For this purpose, a special tool was prepared, essentially consisting of two C-shaped housings, that allowed them to close around the tube and straighten any folds through a closing system. At the same time, this tool also allowed us to test the insertion of the tubes and mark the right length for the point of insertion in the Fitting. It was astronaut Chris Cassidy who suggested marking that point with Kapton tape, previously prepared inside the station by Drew and me, thus creating a reliable reference at the time of junction of the pipes.

Then the time came for the connection of the new cooling system: the tubes are numbered, they all carry a label to help the astronauts insert them in the Fitting for an optimal length. All the Fittings prepared by the team on the ground, and tested among others by Ken Bollweg, were ready to play their fundamental role of swaging and leak-checking.

Thanks to the presence of two decoder rings in the Fitting, I could use the hexagonal wrench to control the tightening angle. With one hand, I held the Fitting through a knob, while with the other I applied the torque closing with the 7/8" hexagonal wrench. We already knew the last two pipes on the lower part of AMS-02 were the most problematic ones, due to the difficulty of access; we were well aware that their straightening would not be easy at all. We had to work a little, but then we managed to straighten them, plug them into the AAF system and plug them into the new system. Once the joints were completed, we had to cover the pipes with layers of MLI to ensure good thermal insulation from the outside environmental conditions. This was the last task of that EVA.

7 The Closing Act

Onboard the ISS, in the weeks after the first EVAs, various activities had taken place and other space vehicles had arrived. On January 25, 2020 [8], about two months after the third spacewalk, Luca and Drew were ready to complete the work. A very tense moment was waiting for them out there.

The name of the activity intended to reflect our main task for that day: closing AMS-02 with the MLI insulation blankets. There weren't any particular risks and with the mission's ground control we decided that it could be a good opportunity for Drew to act as EVA1,

that is the person in charge of the operation. We were quite happy for Drew, because this change of roles would have meant not only being in charge of the EVA, but also for the first time being attached to the CanadArm2, certainly an important experience to have. As I was not going to ride the robotic arm, I positioned myself on the work area where 6 out of 8 connected tubes were reachable. We had shared the tasks, and decided that Drew would check the two tubes positioned in the lower part of AMS-02.

In the meantime, a reasonable quantity of CO_2 was taken by the AMS-02 tank to pressurize the pipes, new connections included. This pressurization would have allowed the astronauts to check the tightness of the fittings installed.

I agreed with mission control that the first pipe I would check was number 5. While I was trying to open the lever that enclosed the piston of the VLI (the Visual Leak Indicator), I immediately realized that something was wrong: that lever had to open quite easily, almost without applying any force and instead, I struggled a lot to open it. Once opened, I clearly saw the red band, which meant that the connector inside was leaking.

The first check carried out by Luca was already showing a problem, and there were seven more to check! All the other tasks performed during the previous EVAs and their long preparation depended on this very moment. Despite the great preparation, the disappointment could be perceived in Luca's voice as he communicated with mission control: "*I have red showing on number five.*" In the control room in Houston and at the POCC at CERN, the tension rose, the colour and the expression on the faces of the operators and physicists of the AMS collaboration changed, the worst possible scenarios started to cross their minds (Figs. 4 and 5).

At mission control in Houston, Ken Bollweg was wondering if he had omitted any details, if something had not been carefully considered in the preparation of the procedure. The first connector was leaking and, even if the subsequent actions were carefully prepared, the team could not stop wondering what went wrong with that connector and what would happen with the others. "*You can imagine, the first*

Fig. 4 Looking up from the POCC at CERN to the giant screen showing the live streaming of the 4th and final Extra Vehicular Activity for repair on the AMS detector. (Credit CERN-PHOTO-202001-021-13)

Fig. 5 Physicists of the AMS Collaboration in the POCC at CERN (CERN PHOTO-202001-021-15)

connector had a big leak, what was wrong? Did we make a mistake? Is there something wrong with the visual leak indicator? Was any calculation wrong? All these questions were in my mind, and of course we were prepared to face situations like this and a series of procedures were ready to deal with this one, but during all of them I was really sweating bullets."

Outside the ISS Luca was facing a stressful moment, it was important to put in place everything necessary to solve the problem or at least try to solve it.

For an astronaut, the execution of the work is the greatest testimony of the preparation and commitment put in place to achieve a result, therefore it was a great disappointment to be dealing with a problem at the first connector, without knowing exactly what to expect at that point from all the others. I knew exactly what I was supposed to do next, which was to use the hex wrench again and tighten the connector once more flat, 1/8 of a turn more (45°). In my mind, the idea of having to cut the pipe and apply a jumper—that is a piece of pipe with two connectors at the ends—was becoming realistic. The time needed for this additional operation also made us think that a fifth EVA would be necessary. I would have left the ISS about ten days later, so there would have been no time for me to be part of it. A fifth EVA would have had a major organizational impact, as it would have involved another team of astronauts.

Luca then let the residual gas out of the joint system and with the hexagonal wrench rotated connector number 5 by another 45 degrees. While waiting for the second pressurization, to check the piston again and verify the presence or absence of the red band, he would continue checking the remaining joints with Drew.

Mission Control: "*Go for number 1 Luca*".

Luca: "*Number 1 is fully open and I have no red*".

Mission Control: "*Ok, number 1 is good*".

We checked all the other connectors and we realized they were all leak-tight. It was great news, the only one with a problem was number 5, the first one I checked, which was also the last one we connected during the previous EVA. After about an hour, I went back to the connector of pipe number 5 for the second check.

Luca opened the connector lever: "*We do have a leak*".

The red band was still there, the connector was still leaking. "*At this point we were in uncharted territory, it had never happened that, after applying one more flat, the leak persisted. From mission control, I received a recommendation that was not foreseen: since we would have to cut this connection anyway to apply the jumper, trying everything and apply a second additional flat was the thing to do. Drew and I invested a lot of time in the preparation of all EVAs, all the details were discussed many times and although in the middle of an unexpected situation, we had to deal with it. So I applied the additional flat and while waiting for the new pressurization of pipe number 5 I quickly went back to the Airlock to collect the required tools and equipment in case we had to install the jumper. In about 25 min, I was already back to AMS-02; Drew and I reversed our roles, I positioned myself again on the robotic arm about 45 min from when I had torqued the AAF a second time. Then the moment to check the VLI again came."*

Luca: "*No red*".

Mission Control: "*No Red?*".

Luca: "*No red!!!*".

At this point a race against time began. Would we be fast enough to install all the MLI covers and blankets required to be able to restart AMS-02 as planned? From mission control they didn't believe there was enough time: this part of the procedure was impossible to test during our training, it was not really possible to simulate it under the same conditions, we could only try to imagine it. With Drew, we actually discussed this part many times, exchanging opinions, ideas and trying to visualise what to expect and what to do. In our heads, we reviewed every step of the installation of these covers and we were able to complete the operation in time. And this is what we did.

While the giant image of Luca Parmitano holding that 7/8" wrench in his hand occupies the giant screen, an explosion of relief, (Fig. 6) gratitude and pure joy pervaded every face in the AMS POCC at CERN. A relieved and as joyful as rarely can be seen (Fig. 7) Sam Ting declared: "*William Gerstenmeier, head of human spaceflight at NASA, thought this it was such a complex thing to do, because it was totally unprecedented. But NASA has to learn to do new things in space, because if they want to have a base on the Moon, or go to Mars, for instance, you need this type of things. We are very grateful to the astronauts who accomplished a fantastic feat, an almost impossible mission.*"

Thanks to that "SIMPLE 7/8" wrench, Parmitano was able to solve a problem that had not occurred during the training in the labs. The astronaut considers the image of him holding "**the right key**" as the symbol of the entire AMS spacewalk programme. "*I liked that photo immediately, because it showed that complex problems are often solved with the ingenuity of the simplest tools such as a lever, first used by Archimedes and represented by that wrench, which allowed us to repair such complex equipment as AMS-02 in this case.*"

Ken Bollweg describes his personal feelings during those moments while he was at Houston control center by recalling the "horrible thoughts" he had when he saw the first fitting checked and leaking. "*I was really shocked and I was trying to understand what could be wrong in the procedure. Then I was a bit relieved when Luca and Drew*

Fig. 6 Luca Parmitano outside the ISS with the wrench and tools in his hand (Photo credit NASA)

tested the rest of them, and one by one they were fine. So, when the other seven were tested fine, I thought that it was just number 5, so we could try to fix it. When I saw the leak again after the additional tightening, I was really astonished and puzzled: we had never seen it before! But then when, after another try, it was fine, all was ok in every sense and the system is OK still now. The leak rate is so low that it is currently undetectable by the methods we have."

Back in the POCC in Geneva, the next two days were busy although not tense at all. Zhan Zhang, the MIT lead engineer of the « Upgraded Tracker Thermal System» recalls [9]: "*After the successful EVA, we knew our cooling system was really tight,*

Fig. 7 Samuel C.C. Ting in the POCC after the successful repair which extended the life to AMS-02 on the ISS (Photo Credit CERN-PHOTO-202001–021-40)

so we had a lot of confidence for the following operation, which we carried out the next day: filling the system with CO_2, so that we could run it. It took 10 h to fill our circulating loop with 1.3 kg CO_2, plus another 6 h to check. On the following Monday, we could finally run the pump. Then we monitored the system for 24 h, before we could start it up for good. And that too was very successful. Finally, we were able to power on the tracker 100% and it started taking the science data. Since then, It has been running 100% smoothly. We also switched the 3 new pumps to check their performance after the launch, and they all behaved normally and smoothly. By noon on Monday, the entire detector started working in full configuration [10,11]." Giovanni Ambrosi, AMS physicist from INFN, also in the POCC, is relieved to say that: "*We are now taking data with the full detector running in nominal conditions and we are really happy, because we now have ahead of us another 10 years and more of good data for science and an increase in the reach of energy where we are able to measure cosmic rays* [12]."

Samuel Ting is confident the detector is entering a new era of discovery, another twenty years of new life. "*We have measured many particles, electrons positrons, protons, antiprotons and all the nuclei, they all have distributions as a function of energy. None of the distributions we measured, agrees with current theoretical model. Everything was measured with extreme precision—excess of positrons, antiprotons, antimatter- nobody had expected. All the distributions with energy of carbon, oxygen, helium, lithium, beryllium, boron nuclei, all their distributions, the behaviour of their distribution with energy is totally unexpected. We are entering a region nobody has never been in before* [13]."

Since the day of restart in January 2020, AMS-02 has been working 24/7, the silicon detector is back to its excellent performances, the cooling system is working

wonderfully well and the experiment is collecting data without a glitch. For a few moments, the physicists and engineers who worked on this fantastic repair adventure feared the worst, as the risk that AMS-02 could never be put back into operation had begun to become more and more real. But the careful preparation of the astronauts and the ingenuity of those who developed the tools necessary to work safely in space, made the success of the mission.

Eventually, all it took was the right key at the right moment to give AMS-02 the possibility to continue making science and—who knows—maybe history!

References

1. Cern Press release, The AMS detector heads for international Space station, 27 April (2011). https://home.cern/news/press-release/cern/ams-detector-heads-international-space-station
2. C. Bortolin, The ALICE experiment at CERN: development and commissioning of the cooling and control systems of the Silicon Pixel Detector and optimization of the LHC Exchange Interface project. PhD Thesis Università degli Studi di Udine, Italy. CERN-THESIS-2010–262 (2011)
3. P. Catapano, Video Full interview of prof. Sam Ting (2020). https://videos.cern.ch/record/2707890
4. P. Catapano, Video Full interview of engineer Ken Bollweg (NASA) (2020). https://videos.cern.ch/record/2707888
5. C. Bortolin, La chiave giusta. Convegno ITP CERN e evento PassioneScienza, 4 Giugno 2021 (2011). https://www.youtube.com/watch?v=-_TXDtYdYkg
6. Expedition 61 - EVA 59 - Luca Parmitano and Andrew Morgan repairing the AMS experiment. 31 May 2020 https://www.youtube.com/watch?v=vAJMiZ2xuo0
7. P. Catapano, Video Full interview of engineer Corrado Gargiulo (CERN) (2020). https://videos.cern.ch/record/2707891
8. Spacewalk to Repair Alpha Magnetic Spectrometer Outside International Space Station on Jan. 25, 2020
9. P. Catapano, Video Full interview of engineer Zan Thang (MIT) (2020). https://videos.cern.ch/record/2707893
10. CERN News, A new life for AMS 6 February, (2020). https://home.cern/news/news/experimets/new-life-ams
11. P. Catapano, Video A new life for AMS (2020). https://videos.cern.ch/record/2708974
12. P. Catapano, Video Full interview of physicist Giovanni Ambrosi (INFN) (2020). https://videos.cern.ch/record/2707892
13. Cern Courrier, AMS detector given a new lease of life. 25 March (2020). https://cerncourier.com/a/ams-detector-given-a-new-lease-of-life/

Gravitational Waves

Gravitational Waves: Why and How

Federico Ferrini

1 Introduction

Light travels faster than any other signal, and light radiation has a great chance of manifesting itself. If we cannot reveal it with optical telescopes, then we can use radio telescopes or detectors capable of capturing infrared and ultraviolet light. We have built satellites capturing X-rays and even more intense gamma rays. We collect information by discovering particles that arrive on the Earth's surface with extraordinary violence—cosmic rays—and other ones, very light and very evasive, running almost at the speed of light, neutrinos.

But there is a part of the Universe unattainable by such means of investigation. Indeed both light and particles in their journey can find something that stops and absorbs them and this prevents us from reaching the boundaries of the Universe. Yet there is a traveler without such constraints, who travels at the speed of light and that nothing and no one can stop. Being able to perceive what is beyond the distances reachable by the most powerful gaze is a beautiful challenge and stimulates the imagination.

However, it takes two key ingredients, the first is the theoretical idea that there is a family of signals that goes beyond everything, without getting lost, without camouflaging, without distorting, with a clear and uncontaminated memory. And this idea was given to us by Albert Einstein, always ready to provide us with amazing and visionary ideas, solid as the hardest rock, but ethereal as the lightest and most vibrant dancer. An idea to be verified, as he himself thought, perhaps not verifiable at all [1, 2]. The second ingredient is to understand what the sense is to be used to perceive these signals. Touch allows us to go not far, taste and smell limit our

F. Ferrini (✉)
Cherenkov Telescope Array Observatory, Via Piero Gobetti 93/3, 40129 Bologna, Italy
e-mail: federico.ferrini@cta-observatory.org

European Gravitational Observatory in Cascina (2011–2017), Pisa, Italy

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_14

range of action. The view is the one that allows us to get to the impenetrable wall—the dark age—but not beyond. So what? There is hearing, which we have not yet considered. Listening to the Universe is what the search for gravitational waves aims to do, listen to its whispers, sighs, the most chilling screams tracing the distribution of the most dramatic collapse phenomena but also the subdued buzz that collects millions of dramatic stories now very distant and so distributed during the expansion of the Universe to make them appear like a sea of fireflies, which, however, instead of emitting light, send sweet sounds.

Why have these gravitational waves not been revealed for about a century; after all they carry a lot of energy on their shoulders and have formidable information potential on very particular astronomical situations. The answer will come in the following, but I can already anticipate that the nature of the gravitational signal is subtle and what the waves combine in their journey from the source to us is very peculiar. This determines the great difficulty of signal detection. Moreover, the measurement of the passage of gravitational waves is the most extreme ever in physics, obscured by the interference of many external causes.

The search for gravitational waves has been the scientific adventure of the twentieth century!

2 The Discovery of the Century: A New Gateway to the Cosmos

If it were a film, we could start telling the story of the search for gravitational waves from the end, or at least from one of the endings at our disposal, of the adventure begun a hundred years ago. The protagonists of this first finale will be the interferometers LIGO Hanford, LIGO Livingston and Virgo and, as the *prima donna*: GW150914. It was detected at 09h50:45 UTC on the 14 September 2015 from binary blackholes merging occurring at about 1.3 billion light-years away. The energy released was enormous $3.0 \pm 0.5\ c^2$ solar masses. [3].

Newspapers around the world have classified the first detection of gravitational waves as "the discovery of the century." Five arguments justify this statement:

(1) The hundred years since Einstein's formulation of general relativity and the article in which he predicts the existence of gravitational waves.

(2) The exceptional nature of the discovery that opens a completely new perspective for the investigation of the Universe, comparable to what Galilei did in the seventeenth century the adoption of the telescope instead of the naked eye to aim at the sky. After Galilei it was an interweaving of technological advancement and scientific inventiveness; from then on, the Universe has been studied almost exclusively with the electromagnetic spectrum, in the multiple frequency ranges, from optical to radio, to the X or gamma band, infrared or ultraviolet, but always electromagnetic radiation emitted by stars or similar matter (plus neutrinos). With the discovery of gravitational waves,

we went much further and added a new method of investigation, independent and complementary, and a new sense to perceive the Universe.

(3) The direct view of black holes.

(4) The extraordinary nature of the numbers at stake: duration, energy, frequencies, infinitesimal measures of extraordinary phenomena.

(5) The surprise of what has been observed. Those working in the field expected to see the effect of the collapse of binary neutron star systems, in principle much more frequent than black holes and therefore much more likely as sources.

A century after the formulation of general relativity, nature has manifested itself in its widest generosity and in a few years the Universe will probably appear different from how we have perceived it until today.

3 Why

The concept of "instantaneous remote action" in Newton's theory could not satisfy Einstein. The causality introduced in its special relativity contrasts with Newton's formula according to which any object with mass, regardless of its state of motion and the distance at which it is located, instantly exerts the action of gravitational force everywhere.

The paradigm of the force of gravitation as resulting from the action at a distance between two bodies having mass is modified by Einstein's theory: every single body with mass influences the space–time around it by deforming this "fabric" that fills the universe.

The gravitational force is not transmitted instantaneously at arbitrary distances but as a deformation of space–time and because of the state of motion of the source propagates in a finite time. Gravitational information manifests itself in the form of deformation of space–time at the speed of light.

The fundamental equation of general relativity links the presence of mass (or energy, thanks to their equivalence, again introduced by Einstein: $E = m\,c^2$) —agent cause—with the deformation of the geometry of space–time—direct consequence.

The equation is in descriptive terms:

Deformation of Geometry = (Coupling Constant) * (presence of mass or energy).

The coupling constant weights the intensity of the effect and is such that within the Solar System the Newtonian formula remains valid.

The Einstein formula is hence:

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

$T_{\mu\nu}$ contains all the information related to the presence and distribution of mass/energy,

$G_{\mu\nu}$ describes the deformation of the four-dimensional geometry,

$G = 6{,}67\,10^{-11}\ m^3/kg\ s^2$ is the universal gravitation constant as defined by Newton, the one which determine the gravitational effects also on the Earth surface, c is the speed of light equal to $3 \cdot 10^8$ m/s.

The coefficient has the value of about $2 \cdot 10^{-43}$, a dramatic small number; the lattice of space–time turns out to be very rigid, with a very low elasticity.

Due to the coupling constant, we can immediately deduce that a mass or a concentration of very large energy is necessary to be able to cause changes in the geometry, capable of canceling the attenuation effect that comes from the coefficient.

And here Einstein again amazes; he deduced three effects from his general relativity theory that have been fully verified by astronomy in the decades following their formulation:

(1) Precession of the Mercury orbit of a value about twice as much as Newton's gravitational physics could justify.
(2) Gravitational shift towards red, made evident for example in the observation of the slowdown of clocks immersed in the gravitational field with respect to an external observer.
(3) Gravitational lensing effect, for which even the Sun can deflect light, and for which from the Earth you can receive the light emitted by distant stars, physically hidden from our direct vision by the solar sphere that is on the direct path between us and the star.

Having at his disposal the mathematical formulation that fully satisfied his conception of gravitation, Einstein tried his hand at finding solutions to the same equations, an extremely difficult task due to the complexity. It is in fact a system of equations coupled with non-linear terms, impossible to solve in the general case with analytic solutions. This is not uncommon in physics, indeed unfortunately it is often the rule, so only by making use of powerful electronic computers and settling for approximate solutions can we determine particular solutions of the general equations.

Einstein then sought a solution that could be calculated in the simplified situation, which results from being very far from the area where the mass is concentrated, in order to approximate the effects on the curvature of space–time and be allowed to treat these effects as a small perturbation of empty space–time, free of masses.

This fundamental simplification allowed him to determine the reduced equation that the small perturbation must satisfy; the resulting equation is identical to the one describing the propagation of electromagnetic waves. He determined the solution of this equation: gravitational wave, ripple of space–time generated by a system of high concentration of mass/energy at distances large from the size of the area surrounding the source.

This is Einstein's fourth deduction from his equation for gravitation and unlike the others it took almost a hundred years to be verified experimentally, thus allowing to dissolve the last residual doubt about the validity of Einstein's theory.

4 How

Gravitational waves are not "material" waves such as pressure waves characteristic of sound phenomena whereby the thrust of the initial fluid medium is transmitted with compressions and rarefactions in matter. Nor are these waves like electromagnetic ones. Gravitational waves are connatural to space–time, to this lattice permeating the universe and therefore can only generate effects on space–time itself, they do not move material bodies in their travel, bodies do not undergo accelerations; the waves cause a modification on the immaterial lattice of distances and times: they dilate the spatial dimensions in one of the two directions perpendicular to the propagation and shorten the distances in the other alternately in the succession of the cyclic perturbation.

The signature in space–time generated by the passage of a gravitational wave is therefore the fundamental characteristic that allows us to identify the phenomenon: the waves are not absorbed by any material and cross unscathed Earth, planets, stars, galaxies except for the effect of the reduction of the intensity inversely proportional to distance. The intrinsic reasons for the complexity of detection are:

(1) The peculiarity and rarity of the sources: Binary systems of neutron stars and/or black holes are infrequent and therefore one must be able to cover a large volume of the universe to have a non-negligible probability of capturing one.
(2) The weakness of the signals generated at great distances from the Earth.
(3) The lack of interaction with standard physical detection systems.

All these complexities have contributed to the need for adventurous and reckless exploration along unknown roads.

In the sixties, the American physicist John Weber devised a simple and refined system at the same time: a large suspended and insulated aluminum cylinder could act as an antenna; the oscillations induced by the passage of gravitational waves would have been measured thanks to the piezoelectric effect. The large bar in the beginning would resonate at its own frequency, about 1.000 Hz, due to the release of the energy of the waves emitted by some galactic supernova; the coincidence of the signals recorded by two antennas placed in distant sites would mark the passage of the wave [4, 5].

Too many internal noises, mainly thermal and external, made the system not very sensitive, and other researchers tried to improve and develop Weber's original idea in much more complex forms. Among these was the contribution of Edoardo Amaldi and his collaborators, who built various cryogenic antennas; the cooling to almost absolute zero allowed to significantly reduce the intrinsic noises of the bar [6, 7]. No signals were found during this phase, but in Italy the nucleus of the scientific community was created that would later lead to the realization of Virgo [8–10].

The optical interferometer is an instrument used in physics since the last decades of the nineteenth century. Michelson and Morley [11] measured the speed of light in different frames of reference, a key experiment for Einstein's special relativity.

The equipment measures the difference in length in two perpendicular directions through light. A light source sends a beam on a partially reflective mirror placed at an

angle of 45° to the incident beam, the beam splitter. The transmitted light continues in a straight line towards a mirror, the reflected light goes towards an identical mirror placed in an orthogonal direction. The light rays are reflected from the mirrors and recombine on the semi-transparent mirror; the two overlap and hit a screen, in a direction orthogonal to that of the initial beam. The screen is a sensitive light meter and receives the sum of the two light beams whose waves interfere by adding up or erasing, depending on the difference in distance traveled on their way. In general, the adjustment of the optical system is such that, in case of absolute quiet, the two light beams are perfectly erased, and no light appears on the screen. When one of the mirrors moves, the distances traveled vary and a light signal begins to appear on the screen that is affected by the difference in optical path in the two perpendicular arms. In our experiment it is not important to measure the precise distance traveled, but to be able to have evidence of a variation that occurred in the two directions, which is precisely what is expected to happen during the transit of gravitational waves.

The interferometer is immediately much more promising than the bars by its very nature: the variation of light is extraordinarily easier to measure than an infinitesimal mechanical variation. There is also a much more important potential of principle: the bar is limited to revealing only the resonance frequency, the interferometer can trace a wide spectrum of frequencies, limited only by intrinsic or surrounding noises.

Before going into more details on the technical characteristics and the struggles we had to overcome to arrive at the brilliant result for which the Nobel prize has been awarded in 2017 it is good to remind the reader that a good summary of the various steps from conception, proposal and realisation are reported in LIGO History and Virgo History as described in their respective web sites (see references for the link).

The first studies of the 70s, followed by some prototypes of reduced dimensions (arms 30–40 m long in Garching near Munich and Pasadena in California), were the guide of large-scale projects (arms of the order of a few kilometers) that began to be conceived starting from the mid-80s. In the 90s the National Science Foundation, in the United States, the Italian National Institute of Nuclear Physics and the French Centre National de la Recherche Scientifique in Europe, accepted respectively the proposal of the LIGO collaboration and the Virgo collaboration. In Hanford, near Seattle, and in Livingston, not far from New Orleans, the construction of the two American interferometers began and in Cascina in the province of Pisa that of the Italian-French interferometer. Interferometers came into operation in 2000 in the United States and in 2003 in Italy, since 2007 they worked jointly.

The feasibility of the projects had been demonstrated by 2012 and the infrastructures capacities to host the interferometers have been built up (Figs. 1, 2, 3 and 4).

The sensitivity was not enough to achieve the objective. The amplitude of the signal of a gravitational wave coming from the collapse of a binary system of neutron stars arriving from the next cluster of galaxies of Virgo from the Einstein theory can be evaluated to be $h \approx 10^{-21}$.

The distance between two mirrors in the interferometer when the wave of such amplitude is passing by is $\Delta L \approx h L$, where L is the distance between the two mirrors and ΔL is the displacement to be measured. On the earth's surface we can

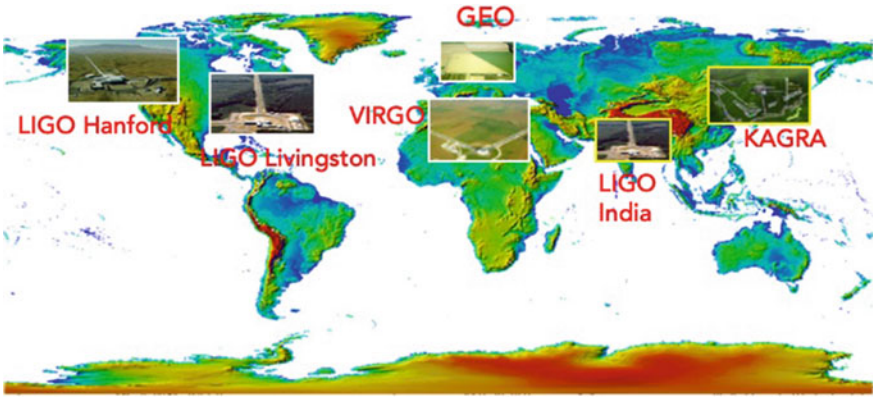

Fig. 1 The Gravitational Waves Observatories in a World-wide Collaboration (Credits The LIGO/Virgo Collaboration)

Fig. 2 Aerial vue of Virgo site (Credits The Virgo Collaboration/N.Baldocchi)

imagine keeping the two mirrors away from each other at distances of the order of the kilometer, so $L \approx 10^3$ m hence: $\Delta L \approx 10^{-18}$ m.

The displacement that the mirror undergoes is equivalent to varying the size of an atom the Earth-Sun distance.

Fig. 3 Aerial vue of LIGO-Livingston site (Credits The LIGO Collaboration) https://www.ligo.caltech.edu/system/avm_image_sqls/binaries/30/jpg_original/ligo-livingston-aerial-02.jpg?1447107179

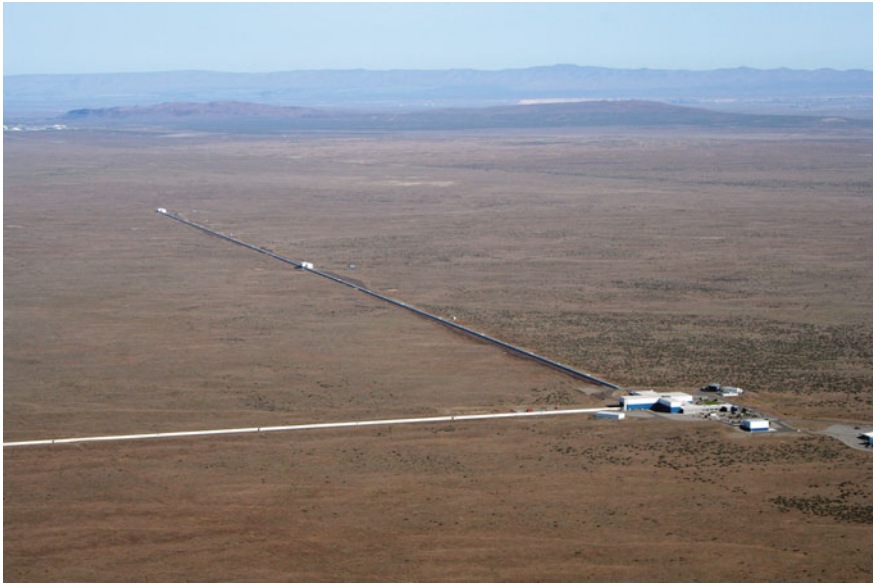

Fig. 4 Aerial vue of LIGO-Hanford site (Credits the LIGO Collaboration) https://www.ligo.caltech.edu/system/avm_image_sqls/binaries/32/jpg_original/ligo-hanford-aerial-04.jpg?1447108890

4.1 An Endless Struggle

4.1.1 Signal Stabilization

The first consideration arising is that the laser must be as stable as possible otherwise its variations could overlap with the optical path difference. The lasers on the market do not have the quality that serves the experiment, in fact they have a possible tolerance of the order of one part in a million, laughing matter compared to the experimental requests at stake. Physicists and engineers of optical groups starting from the best available lasers modify them and create a signal and frequency stabilization system that recreates the output signal from the laser bench, so that it can be injected into the interferometer with the ideal conditions that are required.

The above is a prime example of what we call "noise hunting" because all in all the goal is to remove the intrinsic fluctuations, the noise that the optical signal would carry with it if it were not made pure and very stable.

4.1.2 The Removal of Obstacles

It is also necessary that the light beam does not encounter any obstacles in its path. A speck of dust or a simple molecule would cause the phenomenon known in optics with the name of light *diffusion;* a phenomenon very beautiful to see (it is thanks to the diffusion of sunlight on the upper layers of the atmosphere that the sky appears blue) but harmful to the measurement we want to make and that is already in itself almost impossible to achieve.

The condition of absence of diffusion and scattering of laser light can be satisfied if the beam travels in total vacuum, in the absolute lack of molecules and atoms along its path. This would be the ideal condition, to which we try to get as close as possible by creating the emptiest volume that exists on the earth's surface. In the Virgo Tunnel a volume of 7.000 m^3 is maintained under super high vacuum conditions with a complex system of vacuum pumps (Fig. 5).

The special steel tubes underwent a long and complex treatment which freed their inner walls from the gaseous molecules enclosed in the metal structure. The surfaces of metals naturally capture the molecules that make up the air and trap some of them in their metal lattice. By properly "cooking" the tube sections at high temperatures (400 degrees) and for a sufficiently long time (two weeks) it is actually possible to release the gaseous content, significantly reducing the risk that in the phase of very high vacuum the molecules initially trapped in the metal find it more attractive to migrate to the available empty space, making the work of the pumps much less effective.

The risk of the presence of heavy molecules—for example water, which can be introduced by human bodies even if protected by adequate clothing, or methane, present in the air that returns to the circulation at the time of maintenance operations—which with their relatively large size could disturb the passage of the laser beam,

Fig. 5 The 3 km vacuum tube inside the Virgo North galler (Credits The Virgo Collaboration)

must also be reduced. To this end, a very sophisticated system has been conceived: in four regions of about three meters in size, at the beginning and at the end of the two tunnels, the so-called cryotraps are placed, areas maintained at very cold temperatures thanks to a constant flow of liquid nitrogen, so that the inner face of the pipe has a temperature of about 190° below zero. The heavy molecules are slowed down in their chaotic motion and captured by the inner walls, so as to free their cumbersome presence the path of laser light in the tunnel. Just to give some numbers: the pressure inside the tube is one thousandth of a billionth of an atmosphere (10^{-12} atm).

4.1.3 The Characterization of Optical Components

The material chosen for the optical components is silicon (SiO_2), quartz, 100% pure. It is now very clear to you that we want perfection not for hysteria, but for real experimental needs: the presence of impurities, even very modest and in practice invisible to a standard analysis, produce disturbances on the path of the laser beam causing loss of coherence, diffraction and, if on the surface, aberration.

There are two categories of optical components: the hundreds of lenses, prisms, mirrors of diameters up to 15 cm and the dozen of large lenses and mirrors of diameter equal to or greater than 40 cm and thickness from 15 to 35 cm. For the former ones, there are specialized companies on the market capable of supplying us with these materials in appreciably satisfactory conditions. When the dimensions become significantly larger and the required glass weighs more than 40 kg, the number of highly specialized suppliers is greatly reduced: only three companies worldwide—one in Europe, one in the United States and the third one in Australia—are able to produce the most pure material of the required size. The production of a single

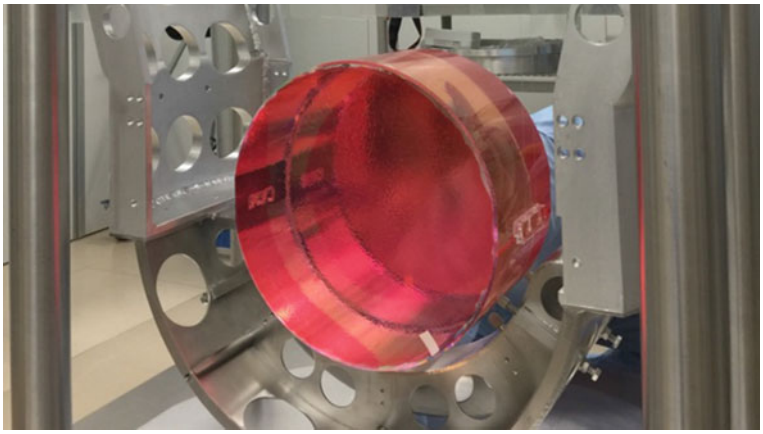

Fig. 6 One of the Virgo mirrors (Credits The Virgo Collaboration/M.Perciballi)

substrate, which will then become a lens or mirror, takes several months, because, even starting from pure material, the cooling times must be very slow, so as not to create microstructures that would then turn out to be critical for the optical quality and the same stability of the sample.

The blocks of super pure glass (at 99.9999%) have, depending on the position where they are to be inserted, perfectly flat surfaces or a spherical curvature; this curvature is calculated through simulations of the overall optical system and results as the portion of a sphere with a radius of about a kilometer and a half, with an accuracy of the order of a few centimeters, to identify with great precision the central part on which the laser beam is engraved (Fig. 6).

At the beginning of its journey in the interferometer, immediately after the complex initial "cleaning", the laser beam is very thin, just a thread of light energy in the infrared that seeks its way to the mirrors at the end of the long arms. Here another brilliant idea intervenes; we have already observed that the longer the path covered by the beam, the better the sensitivity of the instrument will be. The length of the arms is limited by external conditions, for example roads and houses in the Cascina countryside, and, perhaps even more, by financial limitations: an extra kilometer for each arm costs about 20 million euros between infrastructure works and the cost of extending the vacuum system. To overcome this problem, physicists forced the beam to "bounce" numerous times inside each individual arm, practically lengthening the distance traveled. To create this optical cavity, an additional mirror was inserted at the beginning of the arm, with the curved face facing the end of the arm, where the beam is awaited by an identical symmetrically curved mirror. The light thus circulates between the two mirrors several times, obtaining two effects: lengthen the optical path and intensify the beam, which acquires a diameter of about ten centimeters. The amount of light energy inside the vacuum tube becomes significantly higher: in the presence of a laser with a power of 125 W, at the entrance of the central part of

the interferometer the beam has a power of 4.9 kW. With the creation of resonant cavities, the circulating power rises to 650 kW [12].

After about 400 round trips the beam exits and returns to the recombination and measurement zone. A further prolongation of the coming and going would lead to the creation of consistent circulations of light with multiple frequencies than the one we have so well stabilized, and this would damage the definition of the signal. The increase in energy inside the pipe has two effects that we must counteract. The first is the heating effect induced by light radiation on the mirrors: glass, like all materials, if heated, expands and in our mirrors this involves a variation in the radius of curvature. Alas if we gain on one side and lose on the other! We do not dramatize: with an extremely sophisticated system of control of the surface of the mirrors, we calculate in real time their curvature and where necessary we intervene with a fully automated system that injects light from some auxiliary lasers placed near each mirror to bring the surface back to the ideal curvature, heating the vitreous material in appropriate places. It really sounds like science fiction, but it works!

The second effect, caused by the intensification of the beam, is to exert an important pressure on the mirrors themselves, moving them slightly. Light has a dual nature: wave and particle coexist in its properties, and this has been known since the early twentieth century; curiously it is precisely for the interpretation of the photoelectric effect in which the dual nature of light has a crucial role, that Einstein received the Nobel Prize, and not for the theory of relativity! In our beam these properties coexist and photons hitting the mirrors, transfer their momentum. If the surface is rigidly blocked, no variation is observed, but if the surface is free, it will move: so our mirrors move affected by the collective effect of the beam, populated by many photons, as if they were subjected to a total pressure, the radiation pressure. Obviously, a complex control system corrects this effect, well calculable with beam modelling (Fig. 7).

A further consequence of the passage of the beam of substantial dimensions in the optical system is the presence of diffused light. Some photons, due to slight fluctuations of the optical path, leave the parallelism of the beam and can go along trajectories that lead it to bounce on the inner walls of the vacuum tube. A sophisticated system of absorbent rings distributed in the tube and near the main mirrors captures these unruly photons and reduces as much as possible the annoying diffused light that results.

So far, we had to deal with the conditions that we must impose on the quality of the materials that make up the optical components, but we have not yet talked about their surface. A surface defect is equivalent to a noise on the signal and therefore makes it impossible to carry out our measurement. The surfaces of lenses and mirrors must be perfectly smooth and, depending on their use in the optical path, perfectly transparent or perfectly reflective. It is a matter of creating of a perfectly reflective layer for the laser frequency. If for small optical components these characteristics are obtained in a standard way and with performance acceptable by various industries, for large diameter glass there are only two companies capable of standardizing the surface but not of creating a perfectly reflective layer. The only possible solution was to create an ad hoc laboratory, building the appropriate machinery for the treatment of large surfaces, to satisfy scientific requests. The *Laboratoire des Materiaux Avancés*

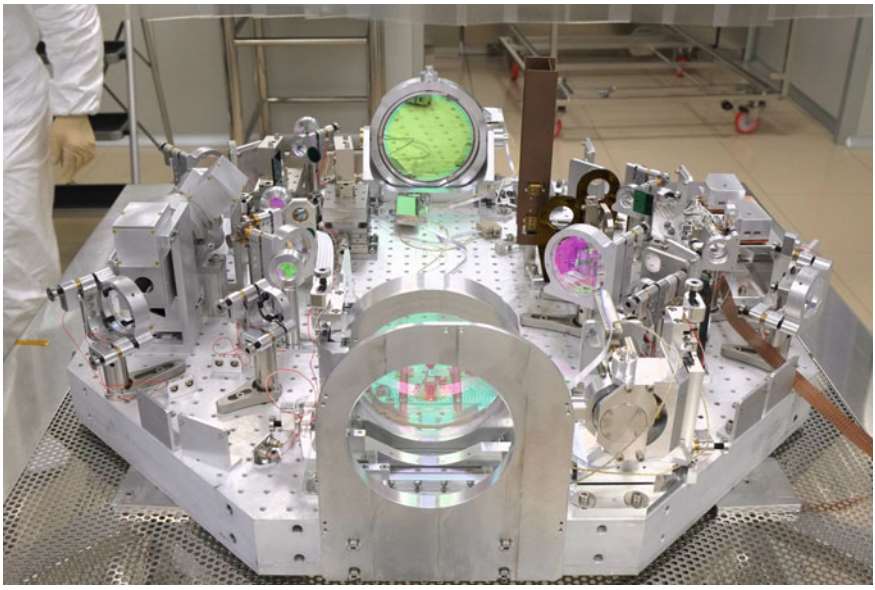

Fig. 7 One of the Virgo optical benches (Credits The Virgo Collaboration/LAPP)

(LMA) in Lyon has been created and the surfaces of large mirrors are coated with combinations of heavy metal oxides with an accuracy of half a nanometer. The laboratory works so well that he was entrusted with the preparation of all the mirrors of the large interferometers, the two LIGO, Virgo and the Japanese KAGRA.

For a complete technical overview with the most recent updates the specialists might consult The Virgo Physics Book: OPTICS and related TOPICS [13].

4.1.4 Suspension and Isolation of Mirrors

The purpose of the interferometer is to measure variations in distances and therefore the position of the mirrors must be known with the utmost precision and must remain immutable all the time, except when gravitational waves pass. In fact, it is the mirrors that determine the lengths of the optical paths traveled by light and modified by the passage of the wave.

Naively you could think of blocking every mirror on a very rigid table well anchored to the ground, but we would be in huge trouble: the earth's surface never stands still, it is subject to a continuous and annoying micro seismicity that generates in the mirrors anchored to it displacements ten billion times greater than the displacement induced by gravitational waves. A deafening noise for our poor measures!

Given the smallness of the expected signal—let's remember that the typical variation is of the order of 10^{-18} m, a billionth of a billionth of a meter—the ideal condition

would be to have the mirrors suspended in the void, without bonds and isolated from everything, free to hear the gravitational rustle and react only to it. Not yet having available mechanisms of levitation of objects of about forty kilograms, we must resort to a system of mechanical suspension, isolating them from external disorders.

This is one of the most complex and difficult problems to solve, to which Adalberto Giazotto, Virgo's father, has managed to give an elegant and very effective solution, devising a brilliant system to super-cushion the vibrations of the ground. Three metal columns 10 m high are fixed to a base on the ground and at the top support a platform, in the center of which are suspended in sequence seven oscillating bodies connected to each other by sections of steel cable, a sequential series of pendulums. Each of these massive objects, called filters, has a cylindrical shape, a diameter of about 70 cm, a height of about 40 cm and a weight varying between 30 and 100 kg, depending on the metal structure and the instrumentation that characterizes it. Each component behaves like a pendulum, capable of swinging in horizontal directions but also of moving along the vertical; each filter therefore has three degrees of freedom.

Twelve blades of special steel are inserted in the lower part of each filter capable of dampening the vibrations coming from the upper layers by a factor of one hundred; in the center of the filter are inserted magnets that, touching each other, couple their appropriately oriented magnetic fields and act as a brake for the vertical oscillations of the system, contributing to the effectiveness of the reduction of external disturbances (Fig. 8).

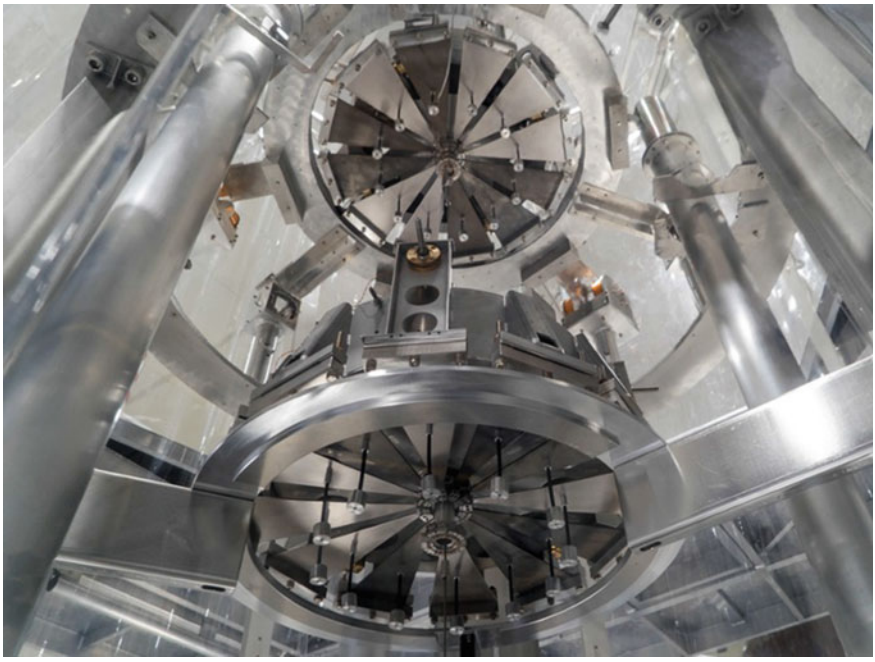

Fig. 8 Virgo Superattenuator (Credits G. Raffaelli/Sonic Somatic)

The last stage of the chain consists of a metal structure reminiscent of the system of wooden sticks used to command the puppets, governed and in their funny movements by a barbell to which they are attached and with light cords. By analogy, the block in which sophisticated magnetic control systems are inserted that allow the mirror to be oriented, in such a way that the alignment of the centers of the mirrors is perfectly in line with the laser beam, is called a *marionette.*

It remains to clarify the way in which the mirror is connected to the marionette. In the first phase of the experiments, thin steel wires were used which, passing under the mirrors, allowed them to be suspended with tranquility and safety, even considering their weight, of the order of 40 kg. The suspension system and the positioning through the marionette are optimally recorded when the towers of the attenuator-mirror complex are in the air, in the absence of the laser beam, to allow the experts to act even physically on the various components. After vacuuming and sending the laser beam on the mirrors, a further adjustment is necessary, through the complex electronic control system designed ad *hoc;* this is necessary due to the temperature variation determined by the presence of the beam. During prolonged phases of operation that can last months, the temperature variations of the external environment modify the internal thermal conditions affecting the materials inside the instrument. Each material, as the temperature changes, undergoes deformations, different depending on the nature of the substance, in particular the responses of steel and glass are quite different. Consequently, to improve the behavior of the mirror-marionette system, it was decided to support the mirrors with delicate quartz threads, hooked to the side walls of the mirrors themselves, excluding any other material than glass. It is not easy at all: just think that each mirror weighs more than forty kilograms and the four quartz fibers that support it have a diameter of a few millimeters!

In conclusion, each suspension tower of the mirrors has seven filters in series, plus three oscillating arms that support everything and help dampen the micro-seismic vibrations of the ground, so that the eleven-meter-high complex produces a reduction in seismic noise of one million billion times, 10^{-15} guaranteeing the mirror an effective isolation from the motions of the ground below [14].

What if there is an earthquake? The answer is simple: we measure its effects thanks to the consequences on the position of the mirrors. The complex attenuation systems could promise interesting applications for understanding seismic motions.

4.1.5 Anthropic Noise

Another problem to be addressed concerns vibrations not due to natural causes but due to human activities, the so-called anthropogenic noise. There are indeed numerous sources capable of generating disturbances to the equipment, which are not always easy to list. Among those well-defined I cite as examples the car traffic, in particular the passage of heavy vehicles, the presence of planes and helicopters that fly over the buildings that house the instrumentation, the rotation of the wind turbines, the presence of quarries and other industrial installations that generate vibrations by crushing and compaction of material. Each of these activities generates superficial or

deep vibrations which distort the measurement process. For example, wind turbines cause low-frequency waves that the rigid layers in the subsoil transmit to the towers that suspend the mirrors, creating a noise around 5 Hz.

4.1.6 Other Sources of Noise

There are other complex sources of noise that limit the performance of the interferometer, and that are related to the intrinsic nature of the electronic equipment that allows us to collect signals: the noise due to the thermal agitation of the conduction electrons and the granular noise due to the corpuscular—quantum—nature of the electrons themselves. They are highly technological noises that we can only fight up to a certain point; one way to protect ourselves is to increase the power of the laser as much as possible, so the effects of fluctuation due to the quantum nature are relatively weaker and the situation improves. Unfortunately, increasing the power of the laser involves risks such as the control of the thermal situation of the mirrors, and therefore it is necessary to find a compromise, an acceptable cocktail in order to guarantee an adequate and stable sensitivity to our instrumentation.

It is important to know on which frequency ranges the various sources of noise exert their influence; with theoretical reasoning and with tests and experimental attempts we have understood that seismic effects limit sensitivity in areas of frequency below 2–3 Hz, thermal effects are the most important in areas between 10 and 300 Hz, while quantum electronic effects are the reason for maximum concern above 300 Hz. The brilliant solutions put in place over the years to limit damage are currently ineffective below 5 Hz and above 1.000 Hz; this is the frequency range in which the detectors placed on the Earth's surface have the best sensitivity range. Consequently, we can identify which astronomical sources are candidates to be recognized in this range. Our ideal candidates as gravitational wave generators will be those systems that emit gravitational waves with effectiveness in the frequency range in which our detectors show the maximum sensitivity—between 30 and 500 Hz—: that is, supernovae, binary systems of neutron stars and binary systems of black holes with masses up to about 100 times the mass of the Sun. The rest of the astronomical emitters for now escapes our hunt.

At the end of this listing of the relentless struggle aimed at bringing out of a sea of multiple noises the faint voice of gravitational waves, I quote some "curious" effects that have a disturbing role, in some cases systematic, and for which ad *hoc* solutions have been found. Both marine and terrestrial tides induced by the motion of revolution of the Moon around the Earth have a well-regular and well-known course; they obviously cause shifts of the mirrors that can be corrected equally systematically with an appropriate system of corrections installed on the platforms placed at the top of the columns of super-attenuators: the positioning of the whole chain is modified to cancel the effect of the tides.

The curvature of the Earth, already 3 km away, causes a lack of parallelism between the vertical directions of the mirrors: they point towards the center of the Earth, due to the local acceleration of gravity, and the faces are not correctly aligned. So, the

system at the top of the column of filters must also think about this, to put them back well aligned so as not to make the incident beam wander.

Even non-systematic environmental phenomena, but linked to local weather variations, affect the stability of the very delicate system: intense winds with their vibrations on the external walls of the buildings and the tunnel that contains the ultra-high vacuum pipe, as well as the very intense wave motion on the seashore about a few kilometers from the Virgo infrastructure transmitted through the ground, they can even lose alignment and force operators to spend a few hours resetting the system to recover the laser beam that no longer sees the center of the mirrors. It is therefore understandable why the long blue tube is not flanked by an elegant row of cypresses that would be perfectly suited to the Tuscan landscape but could cause boring vibrations to the mechanical system of the interferometer.

5 GW150914 and GW170817

Increasing the sensitivity by ten times means increasing the distance by ten times and the spherical volume to investigate one thousand times; this was achieved with the second generation of interferometers in operation since 2015; thanks to that the statistical expectation was confirmed and it was possible to discover gravitational waves.

September 14, 2015: a binary black holes system released the first even detected gravitational wave signal.

August 17, 2017: a binary neutron stars system released the first gravitational wave signal correlated to a spectacular series of electromagnetic signals, from gamma-rays observed 1, 7s after the arrival of the gravitational wave, followed by optical and infrared observations.

Since then, the two LIGOS and Virgo have been harvesting tens and tens of events, creating a rich collection of cases, significantly enriching the knowledge of the violent Universe. The Japanese interferometer KAGRA then joined them and now we have a network of instrument ready to systematically study the GW sky.

On January 5 2020 **GW200105** and on January 15, 2020 **GW200115** the merging of black hole with neutron star have been detected and the article *Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences* published on the 29 of June 2021 [15]. From this observation and quoting the LIGO Web site announcement: *Having confidently observed two examples of gravitational waves from black holes merging with neutron stars, researchers now estimate that, within one billion light-years of Earth, roughly one such merger happens per month.*

"The detector groups at LIGO, Virgo, and KAGRA are improving their detectors in preparation for the next observing run scheduled to begin in summer 2022," says Brady. *"With the improved sensitivity, we hope to detect merger waves up to once per day and to better measure the properties of black holes and superdense matter that makes up neutron stars."* (https://www.ligo.caltech.edu/news/ligo20210629) (Fig. 8).

Up to december 2021, 90 gravitational waves have been detected by the global three interferometer network LIGO, Virgo, KAGRA. (see LIGO, Virgo and KAGRA web sites).

6 Arts and Science

Now it is a widely diffuse communication tool, but indeed it was in 2011 that we started at the European Gravitational Observatory, the house of Virgo, to create Arts & Science events. It is evident the "spiritual" connection between these two creative manifestations of humanity, and we promoted a series of events, such as artistic exhibitions and theater events that accompanied our scientific progress during those complex and exciting years. It has been a direct way to communicate to the enthusiasm and the expectations to the public, that very faithfully, from the youngest children to the mature public, followed our events, asking for more performances. The first art work organized at EGO were the "Big Bang" monologue by Lucilla Giagnoni, the concert "La Musica e l'Universo" with the Quartetto Elisa who among other pieces of music executed a perfect version of the Quartetto in Do maggiore KV 465 "Dissonanze" by W.A. Mozart, the piece "Copenhagen" by M. Frayn adapted to a young public by the Teatri della Resistenza, the monologue "Marie Curie" by the Teatri della Resistenza and the piece "The Physicists" by F. Durrenmatt played by a group of scientists from La Sapienza University of Rome. After we opened this door, further activities continued in the years.

7 Conclusion

"The GW community has a roadmap", stated Giovanni Losurdo, Virgo spokesperson and one of the authors of the Nature published paper [16].

"*An extraordinary science plan for the next two decades: a variety of existing and planned projects, which promise to continue the scientific revolution that began with the discovery of gravitational waves five years ago. Nature has published this roadmap, noting once again the increasing attention of the wider scientific community to this field. In the paper, we list a number of fundamental questions to be addressed through GW observations, which represent an extremely ambitious scientific programme, able to dramatically widen our knowledge of the cosmos.*" (Virgo web site).

In the first two decades of the millennium, we have done a fundamental step forward in the knowledge of the Universe. While the network of actual interferometers is producing almost every week a new contribution to the patchwork of the gravitational Universe, next generation interferometers are under preparation: LISA in the space, Einstein Telescope and Cosmic Explorer on Earth. These future observatories, if realized, with enlarge the horizon to which Gravitational Waves can be

detected toward the very high redshift values and will extend the frequency range of potential observations, till the low frequency region that is going to be the domain to be explored by the space project LISA, that will magnificently complement the ground earth observatories.

Furthermore, gamma ray instrumentation is going to be improved by order of magnitude to be ready to work in symbiosis and to produce a holistic vision of the extreme astronomical phenomena, via the creation of the Cherenkov Telescope Array Observatory. With more than 100 telescopes on two sites located in the Northern and Southern hemisphere, the Observatory will allow to investigate electromagnetic signals produced at energies inaccessible to currently accelerators and will help scientists to unveil the behavior of ultra-relativistic particles as well provide fundamental information about the dark matter. I had the privilege to be first involved in Gravitational Waves and then into gamma-ray particle-astrophysics, probably the most exciting aspects of modern physics.

Unexpected physics is behind the corner!

Post Scriptum: If you want to know more, many books, documentary, educational videos also intended for non-experts are available in several languages.

A Brief History of LIGO.

https://www.ligo.caltech.edu/system/media_files/binaries/313/original/LIGOHistory.pdf

Virgo History https://www.ego-gw.it/history/

https://www.ligo.caltech.edu/

https://www.virgo-gw.eu/

https://gwcenter.icrr.u-tokyo.ac.jp/en/

https://www.cta-observatory.org

References

1. A. Einstein Naherungsweise Integration der Feldgleichungen der Gravitation. Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte, pp. 688–696 (1916)
2. A. Einstein Über Gravitationswellen. Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte, pp. 154–167 (1918)
3. B.P. Abbott et al., Ligo scientific collaboration and Virgo collaboration: observation of gravitational waves from binary black hole merger. Phys. Rev. Lett. **11**(16), 061102 (2016). Published 11 February 2016
4. J. Weber, Gravitational radiation. Phys. Rev. Lett. **18**, 498 (1967)
5. J. Weber, Gravitational-wave detector events. Phys. Rev. Lett. **20**, 1307 (1968)
6. E. Amaldi et al., Initial operation of the M = 390 kg cryogenic gravitational-wave antenna. Il Nuovo Cimento C **1**, 497–509 (1978). https://doi.org/10.1007/BF02510108
7. E. Amaldi et al., Measurements with a small (M = 24,4 kg) cryogenic gravitational-wave antenna at T=1,5 K. Il Nuovo Cimento 07–08 **1**(4), 341–359 (1978). https://doi.org/10.1007/bf02525046
8. E. Amaldi et al., Preliminary results on the operation of a 2270 kg cryogenic gravitational wave antenna with a resonant capacitive transducer and a D.C. squid amplifier. Nuovo Cimento C **9**, 829–845 (1986). https://doi.org/10.1007/BF02558082

9. E. Amaldi et al., Sensitivity of the Roma gravitational wave experiment with the explorer cryogenic resonant antenna operating at 2 K. Europhys. Lett. **12**(1), 5 (2007). https://doi.org/10.1209/0295-5075/12/1/002
10. T. Accadia et al., Virgo: a laser interferometer to detect gravitational waves. Journal of Instrumentation **7**(3), 1–119 (2012). JINST 7 P03012 https://iopscience.iop.org/article/https://doi.org/10.1088/1748-0221/7/03/P03012
11. A.A. Michelson, E.W. Morley, On the relative motion of the earth and the luminiferous ether. Am. J. Sci. **34**(203), 333–345 (1887). https://doi.org/10.1007/bf0252504610.2475/ajs.s3-34.203.33
12. F. Acernese et al., Measurement of the optical parameters of the Virgo interferometer. Appl. Opt. **46**, 3466–3484 (2007). http://hal.in2p3.fr/in2p3-00113562
13. V.Y. Vinet, The virgo physics book : OPTICS and related TOPICS (2020)ss. 1–456 First release Feb. 2001. last update September 9, 2020.
14. S. Braccini et al., An improvement in the VIRGO super attenuator for interferometric detection of gravitational waves: the use of a magnetic antispring (1993)
15. R. Abbott, T.D. Abbott, S. Abraham et al., Observation of gravitational waves from two neutron star-black hole coalescences. Astrophys. J. Lett. **915**(L5), 1–24 (2021). https://doi.org/10.1007/bf0252504610.3847/2041-82131/ac082e
16. M. Bailes et al., Gravitational/wave physics and astronomy in the 2020s and 2030s. Nat. Rev. Phys. **3**, 344–366 (2021)

Underground Searches of Dark Matter

Under the Gran Sasso

Cristiano Galbiati and Walter M. Bonivento

1 Introduction

The *Laboratori Nazionali del Gran Sasso* (LNGS) near L'Aquila, Italy, is one of the major underground laboratories in the world. It was founded in 1988 by Prof. Antonino Zichichi, then President of the Italian *Istituto Nazionale di Fisica Nucleare*, and is one of the four national laboratories of the Institute. The core of the laboratory is three very spacious—$20 \times 20 \times 100\,\mathrm{m}^3$—halls, excavated inside the Gran Sasso mountain, at a depth of 1400 m below the peak, connected by service tunnels and easily accessible through the highway tunnel of the Rome-Teramo highway (Fig. 1). The rock overburden provides a shield equivalent to 3600 m of water, sufficient to guarantee a very significant reduction in the cosmic ray flux.

LNGS played a very important role from the 90s through the 10s in the exploration of the physics of neutrinos and their oscillations, hosting experiments such as MACRO [1], GALLEX [2], GNO [3], Borexino [4], ICARUS [5], and OPERA [6]. Nowadays, LNGS is a major hub for the development of new technologies and experiments for the discovery of dark matter through direct detection of weakly interacting massive particles (WIMPs) scatters on target nuclei [7] and the discovery of Majorana neutrinos through the observation of neutrino-less double beta decay [8]. The discovery of dark matter and Majorana neutrinos would be transformational events, possibly opening the doors to the understanding of the dark sector of the Universe.

These important achievements require an enormous effort: a discovery rests on the observation of very rare, extremely low energy events, a handful in an exposure of

C. Galbiati (✉)
Princeton University, Princeton, NJ 08544, USA
e-mail: galbiati@princeton.edu

Gran Sasso Science Institute, l'Aquila, Italy

W. M. Bonivento
Istituto Nazionale di Fisica Nucleare, Sezione di Cagliari, Cagliari, Italy
e-mail: walter.bonivento@ca.infn.it

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_15

Fig. 1 Aerial view of the LNGS. Credits: TQB1, CC BY-SA 3.0 https://creativecommons.org/licenses/by-sa/3.0, via Wikimedia Commons

several tens of tonnes-year. A necessary pre-requisite for a discovery is to shield the experiments from any possible interfering ionization due to cosmic rays, radioactive backgrounds from the detector materials, and the surrounding environment. LNGS offers the best premises to carry out this important science program. First, its spacious underground hall offers the possibility to outfit experiments with sophisticated passive or active veto systems, surrounding the active detectors from every side. Second, LNGS is host to the largest community of scientists working on searches for ultra-rare events, and is therefore a natural hub for the exchange and development of new ideas and breakthrough technologies. Third, LNGS can offer the best services for screening and assay of materials among all underground laboratories, as well as a variety of support services for the design of detectors and ancillary systems.

Today, LNGS hosts dark matter experiments based on three different types of detection technologies: scintillating crystals, scintillating bolometers, and time projection chambers with noble liquid targets (xenon or argon). All these experiments have world-leading sensitivities in their corresponding regions of mass sensitivity. One experiment, DAMA/LIBRA [9], reported observations of signals attributed to dark matter, which have not been confirmed by other experiments. The detectors in operation today at LNGS, thanks to technological improvements and breakthroughs over a span of twenty plus years of activity, lead the charge in size and sensitivity, and are at the forefront of the crucial exploration of this interesting and intriguing side of our universe.

2 WIMP Dark Matter

A body of astronomical observations, including galactic rotation curves, strong gravitational lensing, and the anisotropy of the cosmic microwave background, indicates that known matter only accounts for a small fraction of the mass-energy content of the Universe. The evidence is widely interpreted to imply that the missing dark matter is composed of as yet unidentified elementary particles [10, 11]. Direct detection of such particles would be an extraordinary discovery, opening a gateway to the investigation of the dark Universe and to the understanding of physics beyond the Standard Model. One of the best motivated dark matter candidates is the WIMP, a thermal relic of the Big Bang, which is expected to have a sub-electroweak scale interaction cross-section and a mass in the GeV/ c^2–TeV/c^2 range. The WIMP dark matter hypothesis is also motivated by the ubiquity of WIMP-like particles in extensions to the Standard Model.

3 The Direct Detection of WIMPs

Direct detection experiments can discover dark matter in the form of WIMPs by detecting their scatters on target nuclei in ultra-low background detectors, especially designed to contain and reject background originating from minimum ionizing radiation in favor of the unique identification of the signal from nuclear recoils. The possibility of discovering dark matter through this avenue received a very significant boost in 1985 when Goodman and Witten [7] discovered one crucial feature of the potential dark matter interactions with regular matter.

The energy of a dark matter particle scattering on a target on Earth depends on several factors: the velocity of the dark matter particle with respect to the Earth's rest frame, the mass of the dark matter particle, and the mass of the target particle. In a first approximation (the so-called 'standard halo model'), the dark matter halo surrounding our galaxy, the Milky Way, is supposed to be a non-rotating sphere, static in the galactic rest-frame, with an isotropic and Maxwellian velocity distribution characterized by a velocity dispersion of about 200 km/s. The overall dark matter density is about 0.3 GeV/c^2cm^3. The Earth moves along with the solar system across the galaxy in the direction of the Cygnus constellation at a velocity of 230 km/s, and has an orbital velocity of 30 km/s around the Sun.

As a consequence, the velocity of approach of a dark matter particle to a target is of the order of a few hundred kilometers per second, a non-relativistic velocity approximately equal to one part per thousand of the speed of light. Due to this fact, any interaction of dark matter particles with regular matter is guaranteed to produce an extremely low energy scatter. Indeed, we can expect that scatters on nuclei may produce energies in the range of a few tens of keV. This is still an extremely low quantity, but right in the middle of an especially interesting range. In fact, the De Broglie wavelength corresponding to the momentum transfer necessary

to induce a nuclear recoil of the order of a few tens of keV is of the order of a Fermi, which is the characteristic size scale of a nucleus! Goodman and Witten first realized that this kinematic condition, a direct result of the velocity distribution of dark matter in the galaxy and of its expected mass range, would induce a coherent interaction between dark matter and the ensemble of nucleons in the target nucleus. In other words, it is expected that under normal conditions, the cross section for the interaction of dark matter with the target nucleus may be enhanced by a factor A^2, where A is the atomic number of the target. This cross section enhancement significantly increases the chance of detection of dark matter: so, the publication of the paper by Goodman and Witten in 1985 gave spur to an enormous interest in developing experimental techniques for the isolation of signals from nuclear recoils in low-background detectors, and is generally credited as the start of the race for the detection of dark matter.

The most popular detector targets explored so far are noble elements, chiefly argon and xenon, either in gas or liquid form, and crystals such as germanium, silicon, thallium doped sodium iodide (Na(Tl)), and calcium tungstate ($CaWO_4$). The energy released by dark matter scatters can be measured in the form of heat (phonons), ionization (electrons) or scintillation (photons), or in a combination of the three, depending on the phase and characteristics of the dark matter detector target. Since 1985, direct dark matter searches have made incredible progress. Figure 2 shows the current status of searches for spin-independent elastic WIMP-nucleus scattering: the sensitivity of dark matter searches through this channel has improved enormously, covering the range from $10^{-36}\,\mathrm{cm}^2$ to $10^{-48}\,\mathrm{cm}^2$ in only thirty-plus years! The top of the exploration range, as already hinted above, corresponds to the value of cross sections for weak interactions mediated by the standard electroweak bosons. Through more than three decades of development, thanks to the adoption of innovative schemes for reduction of background and substantial ingenuity, research has pushed the sensitivity in the range of cross section of interest for interactions mediated by heavier carriers that populate interesting extensions of the Standard Model, including especially models based on several Super Symmetric (SUSY) extensions of the Standard Model. As a rule of thumb, improvements in sensitivity in the vertical axis of Fig. 2 are achieved by increasing the exposure, i.e. the product of target mass and time of data-taking; improvements in the lower end of the horizontal axis are made possible by reductions in the detectors' threshold.

The advance in detector sensitivity has been so successful that, only thirty-seven years after the start of the exploration, dark matter searches are already approaching a fundamental limit, set by the onset of nuclear recoils caused by coherent scatters of solar and atmospheric neutrinos on target nuclei through standard weak interactions, the so-called 'neutrino floor' [13] or 'neutrino fog'. In order to push the sensitivity to dark matter beyond this sensitivity level with direct detection techniques, it would be necessary to discriminate nuclear recoils originated by neutrinos from those originated by dark matter, looking at the direction of the nucleus scatters. In fact, as originally noted by Spergel and colleagues [39], due to the relative motion of the Earth and the dark matter halo, the expected distribution of nucleus scatters is highly anisotropic, pointing in the direction of the velocity field of dark matter particles rel-

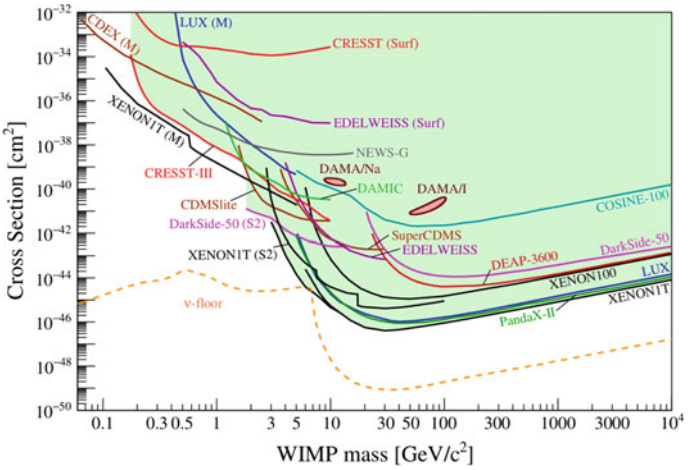

Fig. 2 Current status of searches for spin-independent elastic WIMP-nucleus scattering assuming the standard parameters for an isothermal WIMP halo: $\rho_0 = 0.3\,\mathrm{GeVcm}^{-3}$, $v_0 = 220\,\mathrm{km\,s}^{-1}$, $v_{\mathrm{esc}} = 544\,\mathrm{km\,s}^{-1}$. Results labelled "M" were obtained assuming the Migdal effect [12]. Results labelled "Surf" are from experiments not operated underground. The neutrino-floor shown here for a Ge target is a discovery limit defined as the cross section σ_d at which a given experiment has a 90% probability to detect a WIMP with a scattering cross section $\sigma > \sigma_d$ at $\geq 3\sigma$; it is computed using the assumptions and the methodology described in [13, 14], however, it has been extended to very low DM mass range by assuming an unrealistic 1 meV threshold below 0.8 GeV c^{-2}. Shown are results from CDEX [15], CDMSLite [16], COSINE-100 [17], CRESST [18, 19], DAMA/LIBRA [20] (contours from [21]), DAMIC [22], DarkSide-50 [23, 24], DEAP-3600 [25], EDELWEISS [26, 27], LUX [28, 29], NEWS-G [30], PandaX-II [31], SuperCDMS [32], XENON100 [33] and XENON1T [34–37]. (Note: the figure and the caption are reproduced verbatim from Ref. [38].)

ative to the Earth's reference frame. Technologies for the measurement of the recoil direction are being explored, and in their current state they are still very far from adoption in the very large (tens of tonnes) target required to explore the dark matter frontier sensitivity scale corresponding to the $10^{-48}\,\mathrm{cm}^2$ cross section.

While coherent WIMP-nucleus elastic scattering is the benchmark for WIMP dark matter searches, other kinds of WIMP interactions are also possible and are being thoroughly explored. Target nuclei with non-zero spin could be subject to non-coherent, spin-dependent (SD) interactions with WIMP dark matter (see for example, [40, 41]). Also of interest are possible inelastic scatters of dark matter with target nuclei (see, for example, [42]). In addition, the excellent progress in low-background detection technologies makes it possible nowadays to launch an exploration of certain corners of the direct interaction of WIMP dark matter with electrons (see, for example, [43]).

A convincing discovery of WIMP dark matter will require detection with multiple targets, which will allow for more stringent constraints on dark matter particle properties [44, 45], despite the significant uncertainty in astrophysical parameters [46–48].

4 Key Enabling Technologies

The difference and similarites between the prime high-energy physics laboratory, CERN, and the reference underground astroparticle physics laboratory, LNGS, are striking and telltale.

To face the challenge of building experiments of ever growing size and complexity, both laboratories require gigantic collaborations of several hundreds of scientists, and support the physicists with strong engineering and project management teams, often adopting the most stringent industry standards for construction and commissioning of very large projects. Cryogenics and vacuum technologies, facing the requirement of achieving the lowest temperatures ever experienced in our universe, are key enabling technologies with a strong foothold both in the CERN the accelerator complex and in the underground halls of LNGS.

A subtle issue comes from the need of energy calibration of detectors, as all classes of experiments performed both at CERN and LNGS often require tests with particle beams. Beam tests are particularly crucial for the interpretation and design of the WIMP search experiments performed in underground laboratories, since the low-energy nuclear recoils of interest are characterized by extremely low energy, often featuring a strongly quenched response, which needs careful experimental characterization.

There is, though, a fundamental difference in the enabling technologies: forefront reaserch at CERN Large Hadron Collider (LHC) requires collision of extremely intense beams at the highest center-of-mass energies, producing a huge flow of secondary particles and challenging the detectors on the fronts of data acquisition rates, storage of huge amount of data, ultra-fast detector readout to avoid event pile-up, and radiation hardness. At the CERN LHC, search for rare events, such as the Higgs boson production or possible new SUSY particles, is achieved by digging for gold nuggets with clever analysis methods inside a monumental background of non-interesting events. The challenges in underground astroparticle laboratories are completely different. At LNGS, researchers look for extremely rare processes in nature, triggered not by particle beams but by the natural crossing through detectors of neutral particles wandering in space, resulting in ionizing events at the lowest possible end of detectable signals! The dominant background is often provided by environmental radioactivity originating from natural contaminations in the surrounding rocks and in the detectors' construction materials. Minimizing this is a work-of-art that has its own practices and rules, separated by those *en vogue* at CERN by many decades in ranges of energy and rate. Where the premium focus at CERN is on fast electronic readout and data management, at LNGS maximal appreciation is given to the careful material selection and screening, to the art of designing active and passive shields for the detectors, and to the ever clever schemes for residual background suppression.

5 Solid State Detectors

Two kinds of solid-state detector experiments for DM searches located at LNGS are scintillating crystals, DAMA, SABRE, and scintillating bolometers, CRESST-III and COSINUS.

The DAMA experiment [9] uses 242.5 kg of scintillating NaI(Tl) crystals. This experimental technique does not allow for a distinction between nuclear and electron recoils. Background suppression to WIMP-nucleus scattering relies on shielding and on anti-coincidence technique between signals of different crystals. The experiment claims evidence of an annual modulation of the signal, with a 2.86 t yr exposure, with a significance of more than thirteen standard deviations. The detection threshold improved from 2 to 0.7 keV over time, with the modulating signal lying between 2 and 6 keV. If the signal were interpreted as WIMP-nuclear recoil with spin independent, under the assumption of interactions generated by dark matter distributed according to the standard halo model, the mass of the dark matter particle compatible with such a signal would lie in the range between 10 and 20 GeV c^{-2} with a cross-section of 2×10^{-40} cm^2 on sodium (or between 60 and 100 GeV c^{-2} with a cross-section of 5×10^{-41} cm^2 on iodine).

The DAMA result is very puzzling since, while retaining to date a high statistical significance, it is not confirmed by any other experiment. Within the canonical interpretation of WIMP-nuclear recoils, which is model-dependent and outlined above, the DAMA result is inconsistent with more stringent upper limits from, e.g., XENON1T [34], LUX [49] at the Sanford Underground Laboratory in the US, PandaX-4T [50] at the China Jinping Underground Laboratory, COSINE100 [51] at the Yangyang underground laboratory in South Korea and DarkSide-50 [24]. Among the experiments listed above, COSINE100 also uses the same target, i.e., NaI(Tl) crystals, albeit immersed in a liquid scintillator active veto system, which helps with background reduction. In addition, COSINE100 [52] and ANAIS-112 [53] at the Laboratorio Subterraneo de Canfranc in Spain, another experiment operating NaI(Tl) crystals, both searched for the DAMA evidence by looking for the annual modulation signature: at present, the result of both experiments are inconsistent with the DAMA claim for observation of dark matter with a statistical significance exceeding three standard deviations. Last, but not least, if the DAMA signal were interpreted as due to WIMP-electron scattering, it would be inconsistent with an annual modulation search by XENON100 [54].

New experiments are being designed and prepared to further put to test the claims of DAMA. The SABRE experiments [55] consists of two identical replicas of the same NaI(Tl) crystals setup, to be placed at LNGS and at the Stawell Underground Laboratory in Australia. The purpose of the twin installation is to test the presence of possible seasonal effects in the environmental background. The COSINUS experiment [56] aims at using a set of NaI(Tl) crystals as the target mass of a scintillating bolometer, as discussed in more detail in the next paragraphs.

A bolometer is a device that converts the collected energy into a thermal signal (phonon-mediated). Almost all the released energy is eventually converted into heat,

resulting in an almost non-quenched signal. Scintillating bolometers are a novel type of detectors where the signal from scintillation light of the target crystals is recorded in addition to the classic phonons signal collected in traditional bolometers. The phonons signal reflects the tiny temperature increase of the target crystal following the WIMP scatter. The observation of this tiny signals requires operation of small masses with low heat capacitance at temperatures in the 10 mK range, to sense O (μK) temperature rises.

The CRESST-III experiment [19] uses as a target a 24 g $CaWO_4$ crystal, operating at a 15 mK temperature with a detection threshold of 30 eV. The setup is enclosed in a cryostat equipped with a dilution refrigerator. The detector is operated as a scintillating bolometer. A Transition Edge Sensor (TES) directly in contact with the crystal collects and measured the phonons signal. The crystal is paired with a cryogenic light detector, a silicon-on-sapphire wafer, not in contact with the main crystal, which reads out the scintillation light and is also equipped with a TES. As in other detectors, the ionization and scintillation signals from nuclear recoils are heavily quenched, and the simultaneous observation of the scintillation signal and of the un-quenched total energy deposited (via phonons and heat) allows for discrimination between electron and nuclear recoils. At this time, CRESST-III leads the sensitivity in searches for spin-independent standard WIMP-nucleus scatters below 1.7 GeV c^{-2} (the XENON1T results exploiting the Migdal effect [35] do seem to reach a better sensitivity, but the technique of exploiting the Migdal effect still awaits a complete experimental characterization before gaining full community acceptance).

The same technique of scintillating bolometers will soon be in use with the COSINUS [56] experiment, which aims to probe the DAMA results utilizing a set of NaI(Tl) crystals as its target. The first studies show that the signals extracted at the very low temperature required are different from those of more established bolometer targets and more R&D work is ongoing towards completion of the experiment design.

Outside LNGS, WIMP dark matter is also searched for using other scintillating crystals, such as silicon CCD's, with the DAMIC [22] experiment at SNOLAB in Canada, soon becoming DAMIC-M [57] at the Laboratoire Souterrain de Modane in France, and germanium, with the CDEX [15] experiment at China Jinping Underground Laboratory, and CDMSLite [16] at the Soudan Underground Laboratory and, in assembly phase, SuperCDMS [58] at SNOLAB. Due to the low energy thresholds, down to the single ionization electron, they are sensitive to WIMP masses below the GeV/c^2.

6 Noble Liquid Detectors

Noble liquids, such as liquid argon and liquid xenon, are excellent targets for dark matter searches for many reasons: they can be used to fill large detector volumes to achieve large mass targets sensitive to the smallest cross-sections, they sport high

scintillation and ionization yield, and they are easy to purify, reaching purity levels corresponding to very low electron attachment probabilities, so that the electrons can be drifted for very long distances, and negligible self-absorption of the emitted light.

At LNGS, all detectors based on noble liquids exploit the simultaneous detection of scintillation and ionization signals, using two-phase Time Projection Chambers (TPCs). Ionization events yield both prompt scintillation photons (at 128 nm for argon and 178 nm for xenon), which are detected by light sensors either directly (xenon) or after wavelength shifting to visible light (argon), and a cloud of ionization electrons. The electrons surviving the immediate recombination are drifted towards the anode under the action of an electric drift field uniform throughout the liquid target. Once they reach the surface of separation between the liquid target and the gas pocket near the anode, the electrons are extracted into the gas pocket by a local, stronger extraction field. Within the gas pocket they are accelerated by a local field to induce a secondary, electroluminescence signal, proportional to the number of electrons surviving recombination and drift, and detected by the same light sensors. The scintillation and electroluminescence pulses can be distinguished by their characteristics pulse shapes [59]. In the sole case of argon, the scintillation pulse shape of minimum ionizing events and nuclear recoils is distinctively different, with a strong unbalance between the ns light emitted by singlet dimers (short-lived associations of noble atoms responsible for the emission of scintillation photons) and μs triplet dimers, allowing for an extremely good electron to nucleus recoil separation, up to a factor of up to and above 1×10^8 [23, 60, 61]. In the case of xenon, the potential pulse shape discrimination by scintillation is reset by the lack of difference in the decay time of singlet and triplet dimers, which is always of a few ns. All two-phase detector allow for background discrimination via identification of multiple-sited interactions; by accurate reconstruction of single-sited events and precise selection of a fiducial region; and by measurement of the ratio between scintillation and the (heavily quenched for nuclear recoils) ionization signal, which results both in argon and xenon in a modest, a factor of 100–1000, yet extremely useful background rejection tool.

Noble-liquid detector for dark matter searches have also been built in single-phase, such as DEAP-3600 [25] at SNOLAB and XMASS-I [62] at the Kamioka Observatory in Japan. Compared to double-phase detectors they have a reduced position resolution, i.e., less precise determination of the fiducial volume and of multiple interactions, but they do not not need an electric field and do not suffer from dead time effects due to drift.

6.1 Xenon

Dark matter searches with xenon have been and are performed by various experiments located in three continents, Europe (LNGS, Italy), Asia (Japan and China), and US.

The first liquid xenon TPC installed in LNGS was the XENON10 experiment [63], back in 2006. It contained 15 kg of liquid xenon, followed by XENON100 [33],

with 62 Kg of active mass, in 2008, in the same shield as the XENON10 detector. The passive shield consisted of an outer layer of 20 cm of water, a 20 cm layer of lead, a 20 cm layer of polyethylene, and on the interior a 5 cm copper layer. XENON100 was readout by two planes of 98 and 80 PMTs, respectively.

The world-leading result for high-mass WIMP DM SI searches comes from the XENON1T experiment [34, 36], which was recently disassembled and replaced by XENONnT, now in data-taking. XENON1T was a TPC with 2 t of active mass. The scintillation and electroluminescence light of XENON1T was read out by 248 PMTs. The cryostat housing the TPC was surrounded by a 10 m water tank serving as cosmic veto (Fig. 3). XENON1T reached a sensitivity for WIMP-nucleus SI scattering cross-section of 4×10^{-47} cm^2 at about 30 GeV c^{-2} of mass.

A very intriguing recent result of the XENON1T collaboration is a low energy excess, below 5 keV, over the standard model expectations in the search for WIMP-electron scattering interactions. Though this could be due to residual ^{3}H in the detector, it has also been interpreted as possible solar axion interaction.

The XENONnT experiment [64], beyond the increase of the fiducial target mass to 4t and an almost doubling in the number of PMTs, also included a new neutron veto, filled by ultra-pure water mixed with dissolved Gd salt, to enhance neutron detection efficiency. The experiment is currently in data-taking with the neutron veto filled with water, whereas the gadolinium salt plant is currently under commissioning.

A continuous and successful path towards improving the electron lifetime, i.e., with reducing electronegative contaminants, and suppressing radioactive contami-

Fig. 3 View of the external structure of XENON 1T, experiment devoted to direct search of dark matter, which constitutes 85% of the matter in the Universe. Beside the tank, containing the sensitive part of the detector, it is visible the three levels structure which hosts the apparatus necessary for the functioning of the detector. Credits: ©XENON Collaboration/LNGS-INFN

nants such as ^{222}Rn [65–67] or ^{85} Kr [68], was followed over the years by the XENON collaboration. Indeed, in XENONnT, cryogenic distillation of the recirculating xenon allows to suppress such radioactive contaminants while electronegative impurities are removed by fast liquid recirculation through an oxygen filter. Preliminary results confirm electron lifetimes in the range of those of DarkSide-50, beyond 10 ms.

6.2 Argon

The TPC technique in argon was pioneered by the ICARUS experiment at LNGS [5], a 600 t detector looking for proton decay and beam neutrinos from CERN, and now moved to FNAL again for a search of sterile neutrinos on a short baseline beam [69].

Argon is very abundant in the Earth's atmosphere, about 1% in volume. Commercial production is very abundant as argon is a natural by-product of distillation of air. Ultra-pure argon has a scintillation yield of 40 photons per keV and an electron recoil ionization yield of about 42 electrons per keV for minimum ionizing events, which is quenched to about one-fourth of it for nuclear recoils.

Building on the ICARUS experience, WARP [70] was the first pioneering two-phase TPC looking for WIMP scatters, with a small 2.3 kg detector operated in 2006 and a 140 kg detector operated in 2009.

A few years later came DarkSide-50 [24], again a two-phase TPC with a 50 kg active target read out by 38 cryogenic PMTs (see Fig. 4). The detector was hosted inside a cryostat surrounded by a 30 t boron-loaded organic scintillator neutron veto, located at the center of a 1000 t water Cerenkov cosmic ray veto [23, 71, 72].

Atmospheric argon (AAr) is contaminated by the radioactive isotope ^{39}Ar, a pure β-emitter with an endpoint of 565 keV and a half-life of 269 yr produced by cosmic rays in the outer layers of the atmosphere. Its specific activity in AAr is of about 1.01(8) Bq kg^{-1} [73], which translate to about 1 kBq for a 1 t target. Given that the typical drift time is of the order of 1 ms per meter, the activity of ^{39}Ar in AAr is sufficient to obtain a nearly complete pile-up condition. For this reasons, argon two-phase TPCs at the scale of 1 t and beyond require operation with argon fills strongly depleted in ^{39}Ar. A solution was found by extracting underground argon (UAr) from underground reservoirs in the southwestern corner of Colorado, USA, at a depth of approximately 2000 m gas originating from the Earth's mantle is accessible (the mantle has a significantly lower level of radioactive contaminants than the Crust). The DarkSide Collaboration managed to extract in Colorado over 150 kg of argon, and to purify the entire batch into detector-grad argon thanks to a series of chemical purification units at Fermilab, including a 6 m distillation column. Indeed, ^{39}Ar in the UAr was demonstrated to be very strongly suppressed, with its activity level measured in DarkSide-50 at 0.7 mBq kg^{-1}, almost a factor 1400 below atmospheric argon levels.

In DS-50, argon was continuously purified by recirculating it in gaseous form through a heated getter and a cold charcoal radon trap, reducing contaminants in O_2 and N_2 consistently below sub-ppb levels. The electron lifetime was larger than

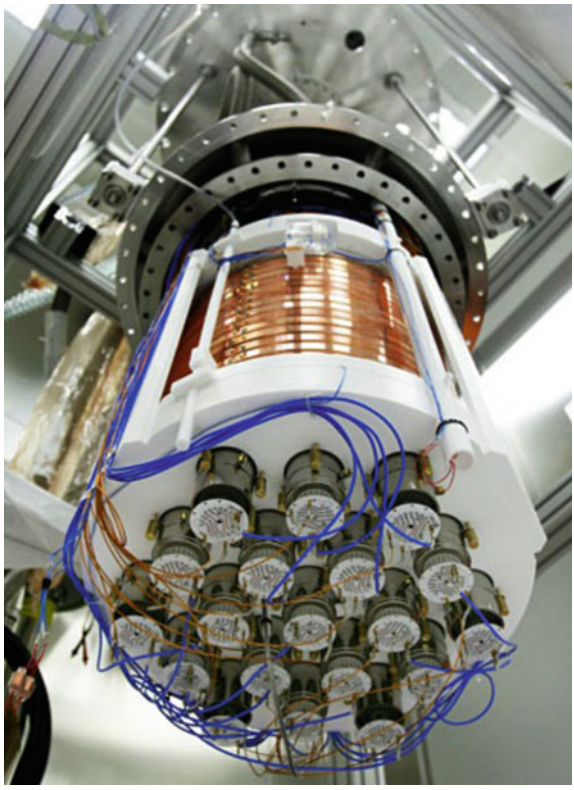

Fig. 4 A view of the DarkSide-50 TPC from below. The scintillation and electroluminescence light emitted by the argon is detected by two planes of 19 PhotoMultipliers. The copper field cage producing the electric drift field is also shown. Credits: The DS-50 Collaboration

8 ms during the whole data-taking, corresponding to 35ppt O_2 equivalent contamination [74]. Thanks to its overall excellent performance, to the outstanding chemical purity of the target and low-background detector, and to the extremely low radioactivity of ^{39}Ar, DarkSide-50 obtained the world best limit for direct dark matter searches in the range of masses between 1.6 and 6.0 GeV c^{-2}. This was obtained by exploiting the lowest possible energy range accessible to the detector, i.e., by observing signals from ionization-only pulses, with the corresponding scintillation signals too low to be of any practical use. The experiment also demonstrated the suitability of searches for high-mass WIMPs with two-phase UAr TPCs, obtaining a null-background result corresponding to a sensitivity of 1×10^{-44} cm^2 at 100 GeV c^{-2}.

Building on the successful experience of DarkSide-50, a new experiment with larger, 20 t fiducial mass, DS-20k [75], was proposed and is now in construction. The DS-20k detector is supported by two technological innovations: UAr extracted with the new Urania high throughput plant in Colorado and purified with the 350 m tall Aria cryogenic distillation column, the tallest ever built; and new cryogenic light detectors based on Silicon PhotoMultipliers (SiPMs).

Argon procurement for DS-20k starts with the Urania plant, whose installation is about to start in Cortez, CO, USA, designed to extract and purify (to 100 ppm) UAr

Fig. 5 A view of the prototype Aria plant for argon purification for the DS-20k experiment. Copyright of the authors

at a rate of 330 kg d^{-1}. The argon collected with Urania in Colorado will be further processed by the Aria cryogenic plant [76], a 350 m tall cryogenic isotopic distillation column, currently under installation phase in a mine shaft operated by the mining company Carbosulcis S.p.A. in Nuraxi-Figus, Sardinia, Italy. The Aria column is designed not only to purify the argon to detector-grade level but also to probe active isotopic separation via distillation for its potential use in a tonne-size detector aimed at improving the world-leading DarkSide-50 result for the search of low-mass dark matter discussed above. A 24 m tall prototype, see Fig. 5, was tested on surface and demonstrated excellent performance, consistent with the design parameters for chemical purification and isotopic separation. As part of the effort on Aria, it is also in the plans the construction of a special detector to measure the activity of ^{39}Ar at the ultra-trace on very small UAr batches, to be put in operation at the Laboratorio Subterráeo de Canfranc in Spain [77].

The scintillation and electroluminescence light readout of DS-20k will consists of two planes, located below the cathode and above the anode, equipped with novel cryogenic light detectors resulting from large-area assemblies of SiPMs equipped with cryogenic amplifiers. The total area covered by SiPMs is of about 25 m^2. The SiPMs, with of 30×30 μm^2 pixel size, have a photon detection efficiency in excess

of 40% at the peak wavelength of 410 nm and, at liquid argon temperature, feature a very small dark noise, below few mHz mm^{-2} as well as a small afterpulse probability, below a few percent.

The two-phase TPC of DS-20k will be built with the top anode and bottom cathode planes made of ultrapure acrylic, and with its sides made of acrylic loaded in gadolinium to boost the signal from neutron captures. The TPC will be contained within a sealed titanium vessel, with the UAr fill between the TPC and the titanium vessel instrumented with light detectors to serve as a neutron veto. The sealed titanium vessel will be in turn hosted inside a protoDUNE-like membrane cryostat [78] with a 600 t liquid atmospheric argon fill also instrumented with light detectors and serving as a cosmic-ray veto.

6.3 Conclusions

WIMP dark matter search with direct detection is one of the major research activities at LNGS. The current and planned experiments aim to maintain the world-leading sensitivity for WIMP-nucleus cross section in both high-mass and low-mass regions.

At low WIMP masses, below 2 GeV c^{-2}, the sensitivity is led by the CRESST-III experiment. At high WIMP masses, it is expected that, the LNGS-based XENONnT experiment and the LZ experiment [79] based in the US Sanford laboratory in

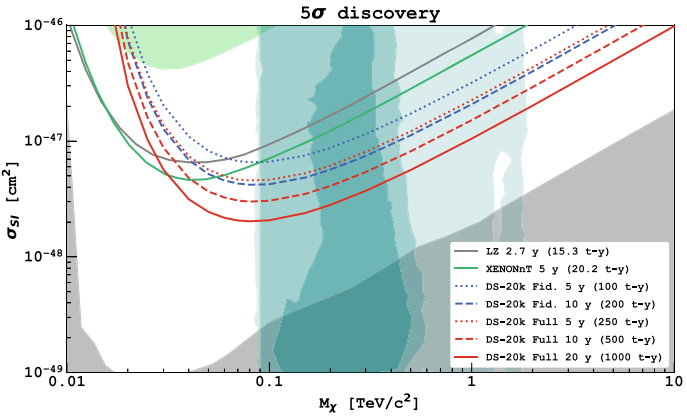

Fig. 6 Projected 5σ discovery potential of DS-20k to spin independent WIMPs for hypothetical different lengths of runs compared to the nominal sensitivity of currently funded experiments LZ (2.7 yr run, 15.3 t yr exposure [79]) and XENONnT (5 yr run, 20.2 t yr exposure [64]) that are expected to lead the field for high mass WIMPs searches in the next few years. The region excluded by XENON1T [36] is shown in shaded green, the neutrino floor for argon in shaded gray [13], and the turquoise filled contours represent the 1-, 2-, and 3-σ favored regions from the pMSSM11 model constrained by astrophysical measurements and the ~36 fbarn LHC data at 13 TeV [82]. Credits: The DS-20k Collaboration

South Dakota, will dominate the sensitivity in the next few years, as shown in Fig. 6. DarkSide-20k results from an initial five year run, expected to begin in 2025, would cover and extend the mass vs. cross-section range being explored by LZ and XENONnT for WIMP masses above 0.1 TeV c^{-2}. Should either XENONnT or LZ measure a dark matter signal over the course of their nominal operating lifetimes of 2.7 yr and 5 yr, respectively, DS-20k is the only planned detector that could promptly confirm their result, and will do so using a different detector technology with different systematic effects. Further plans to extend the discovery sensitivity at large WIMP masses exist with both liquid xenon, the DARWIN [80] experiment, and liquid argon, the ARGO [81] experiment.

Past and recent history teaches us that, at any stage of this race for dark matter, confirmation of a first signal with a different technology would be crucial in establishing the validity of a landmark discovery. Additionally, a positive detection using two different targets will result in improved determination of the dark matter mass and coupling [44, 45].

References

1. M. Ambrosio et al., Measurements of atmospheric muon neutrino oscillations, global analysis of the data collected with MACRO detector. Eur. Phys. J. C **36**, 323 (2004)
2. W. Hampel et al., GALLEX solar neutrino observations: results for GALLEX IV. Phys. Lett. B **447**, 127 (1999)
3. M. Altmann et al., Complete results for five years of GNO solar neutrino observations. Phys. Lett. B **616**, 174 (2005)
4. M. Agostini et al., Identification of the cosmogenic ^{11}C background in large volumes of liquid scintillators with Borexino. Eur. Phys. J. C **81**(12), 1075 (2021)
5. S. Amerio et al., Design, construction and tests of the ICARUS T600 detector. Nucl. Instrum. Meth. A **527**, 329 (2004)
6. N. Agafonova et al., Final results on neutrino oscillation parameters from the OPERA experiment in the CNGS beam. Phys. Rev. D **100**(5), 051301 (2019)
7. M.W. Goodman, E. Witten, Detectability of certain dark-matter candidates. Phys. Rev. D **31**(12), 3059 (1985)
8. W.H. Furry, On transition probabilities in double beta-disintegration. Phys. Rev. **56**, 1184 (1939)
9. R. Bernabei et al., The DAMA project: Achievements, implications and perspectives. Prog. Part. Nucl. Phys. **114**, 103810 (2020)
10. V. Trimble, Existence and nature of dark matter in the universe. Ann. Rev. Astron. Astrophys. **25**(1), 425 (1987)
11. G. Bertone, D. Hooper, History of dark matter. Rev. Mod. Phys. **90**(4), 045002 (2018)
12. M. Ibe, W. Nakano, Y. Shoji, K. Suzuki, Migdal effect in dark matter direct detection experiments. JHEP **03**, 194 (2018)
13. J. Billard, E. Figueroa-Feliciano, L. Strigari, Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments. Phys. Rev. D **89**(2), 023524 (2014)
14. J. Billard, F. Mayet, D. Santos, Assessing the discovery potential of directional detection of dark matter. Phys. Rev. D **85**, 035006 (2012)
15. Z.Z. Liu et al., Constraints on spin-independent nucleus scattering with sub-GeV weakly interacting massive particle dark matter from the CDEX-1B experiment at the China Jinping underground laboratory. Phys. Rev. Lett. **123**(16), 161301 (2019)

16. R. Agnese et al., Search for low-mass dark matter with CDMSlite using a profile likelihood fit. Phys. Rev. D **99**(6), 062001 (2019)
17. G. Adhikari et al., An experiment to search for dark-matter interactions using sodium iodide detectors. Nature **564**(7734), 83–86 (2018)
18. G. Angloher et al., Results on MeV-scale dark matter from a gram-scale cryogenic calorimeter operated above ground. Eur. Phys. J. C **77**(9), 637 (2017)
19. A.H. Abdelhameed et al., First results from the CRESST-III low-mass dark matter program. Phys. Rev. D **100**(10), 102002 (2019)
20. R. Bernabei et al., First results from DAMA/LIBRA and the combined results with DAMA/NaI. Eur. Phys. J. C **56**, 333 (2008)
21. C. Savage, G. Gelmini, P. Gondolo, K. Freese, Compatibility of DAMA/LIBRA dark matter detection with other searches. J. Cosmol. Astropart. Phys. **04**, 010 (2009)
22. A. Aguilar-Arevalo et al., Results on low-mass weakly interacting massive particles from a 11 kg-day target exposure of DAMIC at SNOLAB. Phys. Rev. Lett. **125**, 241803 (2020)
23. P. Agnes et al., DarkSide-50 532-day dark matter search with low-radioactivity Argon. Phys. Rev. D **98**(10), 102006 (2018)
24. P. Agnes et al., Low-mass dark matter search with the DarkSide-50 experiment. Phys. Rev. Lett. **121**(8), 081307 (2018)
25. R. Ajaj et al., Search for dark matter with a 231-day exposure of liquid argon using DEAP-3600 at SNOLAB. Phys. Rev. D **100**(2), 022004 (2019)
26. L. Hehn et al., Improved EDELWEISS-III sensitivity for low-mass WIMPs using a profile likelihood approach. Eur. Phys. J. C **76**(10), 548 (2016)
27. E. Armengaud et al., Searching for low-mass dark matter particles with a massive Ge bolometer operated above-ground. Phys. Rev. D **99**(8), 082003 (2019)
28. D.S. Akerib et al., Results from a search for dark matter in the complete LUX exposure. Phys. Rev. Lett. **118**(2), 021303 (2017)
29. D.S. Akerib et al., Results of a search for Sub-GeV dark matter using 2013 LUX data. Phys. Rev. Lett. **122**(13), 131301 (2019)
30. Q. Arnaud et al., First results from the NEWS-G direct dark matter search experiment at the LSM. Astropart. Phys. **97**, 54 (2018)
31. X. Cui et al., Dark matter results from 54-Ton-Day exposure of PandaX-II experiment. Phys. Rev. Lett. **119**(18), 181302 (2017)
32. R. Agnese et al., Search for low-mass weakly interacting massive particles with SuperCDMS. Phys. Rev. Lett. **112**(24), 241302 (2014)
33. E. Aprile et al., XENON100 dark matter results from a combination of 477 live days. Phys. Rev. D **94**(12), 122001 (2016)
34. E. Aprile et al., Light dark matter search with ionization signals in XENON1T. Phys. Rev. Lett. **123**(25), 251801 (2019)
35. E. Aprile et al., Search for light dark matter interactions enhanced by the Migdal effect or Bremsstrahlung in XENON1T. Phys. Rev. Lett. **123**(24), 241803 (2019)
36. E. Aprile et al., Dark matter search results from a one ton-year exposure of XENON1T. Phys. Rev. Lett. **121**(11), 111302 (2018)
37. E. Aprile et al., Search for coherent elastic scattering of solar ^{8}B neutrinos in the XENON1T dark matter experiment. Phys. Rev. Lett. **126**, 091301 (2021)
38. J. Billard et al., Direct detection of dark matter—APPEC committee report (2021). arXiv:2104.07634
39. D.N. Spergel, Motion of the Earth and the detection of weakly interacting massive particles. Phys. Rev. D **37**(6), 1353 (1988)
40. E. Aprile et al., Constraining the spin-dependent WIMP-nucleon cross sections with XENON1T. Phys. Rev. Lett. **122**(14), 141301 (2019)
41. A.H. Abdelhameed et al., Cryogenic characterization of a $LiAlO_2$ crystal and new results on spin-dependent dark matter interactions with ordinary matter. Eur. Phys. J. C **80**(9), 834 (2020)
42. E. Aprile et al., Search for inelastic scattering of WIMP dark matter in XENON1T. Phys. Rev. D **103**(6), 063028 (2021)

43. P. Agnes et al., Constraints on Sub-GeV dark-matter–electron scattering from the DarkSide-50 experiment. Phys. Rev. Lett. **121**(11), 111303 (2018)
44. M. Pato et al., Complementarity of dark matter direct detection targets. Phys. Rev. D **83**(8), 083505 (2011)
45. J.L. Newstead, T.D. Jacques, L.M. Krauss, J.B. Dent, F. Ferrer, Scientific reach of multiton-scale dark matter direct detection experiments. Phys. Rev. D **88**(7), 076011 (2013)
46. D. Baxter et al., Recommended conventions for reporting results from direct dark matter searches. Eur. Phys. J. C **81**(10), 907 (2021)
47. J.D. Lewin, P.F. Smith, Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil. Astropart. Phys. **6**(1), 87 (1996)
48. C. McCabe, The Earth's velocity for direct detection experiments. J. Cosmol. Astropart. Phys. **02**, 027 (2014)
49. D.S. Akerib et al., Improving sensitivity to low-mass dark matter in LUX using a novel electrode background mitigation technique. Phys. Rev. D **104**(1), 012011 (2021)
50. Y. Meng et al., Dark matter search results from the PandaX-4T commissioning run. arXiv:2107.13438
51. G. Adhikari et al., Strong constraints from COSINE-100 on the DAMA dark matter results using the same sodium iodide target. Sci. Adv. **7**(46), 2699 (2021)
52. G. Adhikari et al., Three-year annual modulation search with COSINE-100. arXiv:2111.08863
53. J. Amaré et al., Annual modulation results from three-year exposure of ANAIS-112. Phys. Rev. D **103**, 102005 (2021)
54. E. Aprile et al., Search for electronic recoil event rate modulation with 4 Years of XENON100 data. Phys. Rev. Lett. **118**(10), 101101 (2017)
55. M. Antonello et al., The SABRE project and the SABRE proof-of-principle. Eur. Phys. J. C **79**(4), 363 (2019)
56. G. Angloher et al., Results from the first cryogenic NaI detector for the COSINUS project. J. Instrum. **12**(11), P11007 (2017)
57. N. Castelló-Mor, DAMIC-M experiment: thick, silicon CCDs to search for light dark matter. Nucl. Instrum. Meth. A **958**, 162933 (2020)
58. W. Rau, SuperCDMS SNOLAB—status and plans. J. Phys. Conf. Ser. **1342**(1), 012077 (2020)
59. P. Agnes et al., Electroluminescence pulse shape and electron diffusion in liquid Argon measured in a dual-phase TPC. Nucl. Instrum. Meth. A **904**, 23 (2018)
60. M.G. Boulay, A. Hime, Technique for direct detection of weakly interacting massive particles using scintillation time discrimination in liquid argon. Astropart. Phys. **25**(3), 179 (2006)
61. P. Adhikari et al., Pulse-shape discrimination against low-energy Ar-39 beta decays in liquid Argon with 4.5 Tonne-Years of DEAP-3600 data. Eur. Phys. J. C **81**(9), 823 (2021)
62. A. Takeda, WIMP search from XMASS-I fiducial volume data with background prediction. J. Phys. Conf. Ser. **1342**(1), 012080 (2020)
63. J. Angle et al., First results from the XENON10 dark matter experiment at the Gran Sasso national laboratory. Phys. Rev. Lett. **100**, 021303 (2008)
64. E. Aprile et al., Projected WIMP sensitivity of the XENONnT dark matter experiment. JCAP **2020**(11), 031 (2020)
65. E. Aprile et al., ^{222}Rn emanation measurements for the XENON1T experiment. Eur. Phys. J. C **81**(4), 337 (2021)
66. E. Aprile et al., Intrinsic backgrounds from Rn and Kr in the XENON100 experiment. Eur. Phys. J. C **78**(2), 132 (2018)
67. E. Aprile et al., Online^{222}Rn removal by cryogenic distillation in the XENON100 experiment. Eur. Phys. J. C **77**(6), 358 (2017)
68. E. Aprile et al., Removing krypton from xenon by cryogenic distillation to the PPQ level. Eur. Phys. J. C **77**(5), 275 (2017)
69. L. Bagby et al., Overhaul and installation of the ICARUS-T600 liquid argon TPC electronics for the FNAL short baseline neutrino program. J. Instrum. **16**(01), P01037 (2021)
70. G. Fiorillo et al., The WArP dark matter search, in *Proceedings of Science, IDM2008* (2008), p. 016

71. D. Acosta-Kane et al., Discovery of underground argon with low level of radioactive ^{39}Ar and possible applications to WIMP dark matter detectors. Nucl. Inst. Meth. A **587**(1), 46 (2008)
72. J. Xu et al., A study of the trace ^{39}Ar content in argon from deep underground sources. Astropart. Phys. **66**, 53 (2015)
73. P. Benetti et al., Measurement of the specific activity of 39Ar in natural Argon. Nucl. Instrum. Methods Phys. Res. A **574**(1), 83 (2007)
74. R. Acciarri et al., Oxygen contamination in liquid argon: combined effects on ionization electron charge and scintillation light. J. Instrum. **5**(05), P05003 (2010)
75. C.E. Aalseth et al., DarkSide-20k: A 20 tonne two-phase LAr TPC for direct dark matter detection at LNGS. Eur. Phys. J. Plus **133**(3), 131 (2018)
76. P. Agnes et al., Separating ^{39}Ar from ^{40}Ar by cryogenic distillation with Aria for dark-matter searches. Eur. Phys. J. C **81**(4), 359 (2021)
77. C.E. Aalseth et al., Design and construction of a new detector to measure ultra-low radioactive-isotope contamination of Argon. J. Instrum. **15**(02), P02024 (2020)
78. B. Abi et al., First results on ProtoDUNE-SP liquid argon time projection chamber performance from a beam test at the CERN neutrino platform. J. Instrum. **15**(12), P12004 (2020)
79. D.S. Akerib et al., Projected WIMP sensitivity of the LUX-ZEPLIN dark matter experiment. Phys. Rev. D **101**(5), 351 (2020)
80. M. Schumann, L. Baudis, L. Bütikofer, A. Kish, M. Selvi, Dark matter sensitivity of multi-ton liquid xenon detectors. JCAP **10**, 016 (2015)
81. C. Galbiati et al., Future dark matter searches with low-radioactivity Argon, in *Input to the European Particle Physics Strategy update 2018–2020* (2018)
82. E. Bagnaschi et al., Likelihood analysis of the pMSSM11 in light of LHC 13-TeV data. Eur. Phys. J. C **78**(3), 87 (2018)

Optical Astronomy

Big Telescopes and Observatories: Hi-Tech Challenges for Great Astronomical Science

Gianni Marconi and Riccardo Scarpa

The most incomprehensible thing about the Universe is that it is understandable.
(A. Einstein)

1 Ancient Astronomy

Humans have always looked at the heavens, wondering about the nature of those little luminous points we call stars. Innumerable hypotheses have been put forward about their nature, with minimal success.

After all, citing Eddington, "*not only the Universe is stranger than we thought, but it is also actually way stranger than we can think of!*".

Thus, the true nature of the stars remained a complete mystery all the way to modern days. Their immutable relative position—it is impressive to think that the constellations we observe today were drawn about four thousand years ago—and complete inaccessibility gave them a divine status.

Moreover, of all the objects visible with the naked eye in the night sky, seven move with respect to the others: the Sun, the Moon, and the five visible planets Mercury, Venus, Mars, Jupiter, and Saturn. These objects became an important integral part of all human cultures, from Mesopotamia to present days, being associated to the most variated and fantastic things. It is not a coincidence that according to the Genesis it took seven days to God to build the Universe, and weeks have seven days named after these moving objects. These fantastic seven appear over and over, and yet it is

G. Marconi (✉)
Alonso de Córdova, European Southern Observatory, Santiago de Cile, 19001 Vitacura Casilla Santiago de Chile, Chile
e-mail: gmarconi@eso.org

R. Scarpa
Department of Astrophysics Research, Gran Telescopio CANARIAS, Instituto de Astrofisica de Canarias (IAC), C/Via Láctea, s/n E-38205, La Laguna Tenerife, Spain
e-mail: riccardo.scarpa@gtc.iac.es

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_16

striking to note how in no culture was it ever proposed the planets are something like the earth and the sun is a star. They were gods, they were magic, they had mystique powers, and even today we go on reading the horoscope.

The earth was thought to be a unique place in the Universe, located right in its center and at rest, created as everything else by God explicitly for us, the humans.

The vast, though still limited understanding we have today of what the Universe really is, took thousands of years to build up, following a tortuous process strictly linked to the technological advances made by our society.

Ancient astronomy was limited to recording the position of stars, mainly for navigation purpose, planets locations within the zodiac, and time tracking—it was Hipparchus of Nicaea around 150 B.C. to introduce the 24 h division of the day and the 365.25 days duration of the year. Attempt were made by Greek philosophers to develop a comprehensive model for the Universe. Unfortunately, these were plagued by fantastic and divine preconceived ideas. By far the most successful one was the geocentric model, the Ptolemaic model, in which everything is believed to rotate around the earth. This model, having no physical bases, was developed according to our everyday experience provided by our senses. No one can feel the rotation of the earth or its motion around the Sun, thus it was logical to imagine the earth was not moving at all. The star's distance was completely unknown at that time, and no one could imagine—this is true even today for many people—how far they are. Claiming they were completing a full revolution around the earth in twenty-four hours was therefore not a problem. Because distant objects appear to move less rapidly that nearby ones, in the geocentric model objects were placed at increasing distance according to their apparent proper motion, the Moon being always considered the nearest among heavenly body, the fixed stars been the most distant one.

A more rational approach to understand the Cosmo was followed by some illuminated minds. The achievement made in ~250 B.C. by Aristarchus of Samos are particularly impressive. He placed the Sun at the center of the Universe and was able to determine the radius of the earth, measure the distance of the Moon and the Sun reaching the conclusion that the sun had to be much larger than the earth. His work and ideas, however, were soon forgotten and Ptolemy's Almagest remained the undisputed truth abouts astronomy until the XVIth century, when Copernicus and Kepler were able to (re)develop the heliocentric model.

Of course, all this was happening in Eurasia. In other parts of the world the situation was different, though in most cases even further from truth. It is worth to remember that Chinese astronomers were regularly recording celestial phenomena, mainly for divinations. Besides recording eclipses and comets, they recorded in 1054 the appearance in the sky of a new star. Located in the constellation of Taurus, the star became so bright to remain visible during daytime for almost a month. Astonishing enough, no trace of this event is to be found in European records. We now know this was the supernova that originated the Crab nebula and knowing the date of the explosion is an extremely important piece of information for modern astrophysicists.

2 The Telescope

For astronomy and humanity as a whole, the real breakthrough arrived in 1609 when Galileo Galilei introduced for the first time the use of the telescope to investigate the sky. Thanks to its revolutionary device Galileo could demonstrate beyond any doubt that the Sun, not the earth, is at the center of the solar system, fully confirming the Copernican ideas, first circulated in 1514. Galileo succeeded to measure the height of the mountains it could see on the Moon, demonstrating they were similar to terrestrial mountains. For the first time in history, humans realized the earth is not special. It is a planet like the others orbiting around the Sun. And this was not the word of Aristarchus of Samos or Copernicus, which could not sustain their claims even if they were correct. This time, for the first time, it was the claim of science, and no one could seriously refute it!

A few decades later, in 1687, Isaac Newton, possibly the most brilliant mind we ever had, put forward in its Principia the laws of dynamics and universal gravitation. With these new theoretical and technical tools in hand, a gold century of discoveries started and by the end of XVIIIth century astronomers had a solid understanding of what the solar system really is. Planetary position and eclipses could be predicted with impressive precision, the distances of the planets were measured. Many other celestial phenomena like comets—up to that moment believed to be atmospheric phenomena—found a rational explanation. Among others, Edmund Halley using Newton's laws was capable to anticipate the passage of a comet (later to be named after him), in this way demonstrating their validity and predicting power.

Finally, in 1838 Friedrich Bessel was able to measure for the first time the distance of a star, 61 Cygni. While the Sun is located at 8.3 light minutes from us, and Neptune at 4.1 light hours, 61 Cygni was found to be at a distance of about 11.4 light years, almost a million times more distant than the Sun. Systematic observations with increasingly powerful telescopes allowed humans to realize, for the first time, that the moon and the planets are bodies similar to the earth and stars are nothing else than distant suns. Today this may seem a silly result and yet, this is a major triumph of science for in no culture anything even remotely close to all this was ever proposed.

To cope with the problem of chromatic aberration inherent to refracting telescopes, Newton made an additional major contribution to astronomy inventing the reflecting telescope, the de facto standard in all modern observatories. Large size reflecting telescopes are easy to build compared to refractors. This allowed the rapid development of powerful instruments of which we recall here the giant 1.8 m built by Lord Ross in 1845. Large telescopes in hand of many brilliant minds resulted in the discoveries and study of many different objects, among other: the planets Uranus and Neptune, many asteroids, comets, other planet's moons, double stars, variable stars, star clusters and nebulae.

3 Extragalactic Astronomy

Galaxies, while routinely observed, were still thought to be nebulae, like the Orion nebula, rather than distant stellar systems. The awareness that we live in a galaxy, the Milky Way, and that there are millions of them, still had to wait for the following major technological breakthrough. This occurred with the invention of photography in 1839. Astronomers, however, had to wait the end of the XIXth century before sufficiently sensitive emulsion became available and immediately applied them to image the night sky. Photography, compared to visual observation, has the double advantage of allowing for large exposure that can record fainter targets and remove subjectivity from observations.

At the beginning of the XXth century astronomy had advanced to an amazing level, and yet we were (and still are!) extremely ignorant. Newtonian dynamics gave to many the impression that the position and velocity of the entire Universe could be predicted. But nothing was known about the physics of stars, their chemical composition, age, and, most important, the ultimate source of energy powering them.

New theoretical tools were necessary to give an answer to these fundamental questions. Help came from a number of brilliant minds that revolutionized both physics and astronomy. The best known of them is Einstein, because his ideas were so advanced we are still fighting to confirm some of them today, but it was clearly the effort of many that brought about the revolution.

Among other things, the photoelectric effect and the black body radiation were understood, paving the way to quantum mechanics. Nuclear power, engraved in the immortal equation $E = mc^2$, was discovered and allowed astronomer to understand how stars can produce the tremendous quantity of energy they emit. Quantum mechanics explained the spectral lines observed in the spectrum of celestial objects allowing the determination of their chemical composition. For instance, the very strong green emission lines observed in the spectrum of many nebulae but never observed in the laboratory, were understood to be forbidden lines produced by oxygen twice ionized [OIII], instead of being the emission of Nebulium, an element suggested to exist only in nebulae.

The time was ripe for a final leap. Edwin Hubble using the Mount Wilson 100 inches telescope, at the time the largest telescope in operation, was able to measure the distance of the Andromeda galaxy, demonstrating it could not be part of the Milky way. Extragalactic astronomy was born. Humanity became conscious that the Universe was immensely bigger than thought before. Moreover, Hubble noticed that the light of distant galaxies is systematically shifted toward longer wavelengths, an effect soon interpreted as due to Doppler shift, indicating distant galaxies are systematically receding from us.

Hubble went on measuring the constant of proportionality between distance and receding velocity, that became known has the Hubble constant.

At that time Einstein was applying his brand-new theory of general relativity to cosmology. Using his own equations, it developed a static state model for the Universe. To avoid galaxies collapsing toward each other under the effect of their

own gravity, Einstein introduced a repulsive term in his equations, baptized the cosmological constant.

As soon as Hubble results on receding galaxies became widely known, Einstein fully embraced the idea that the Universe was expanding and therefore, not needing to avoid galaxies to collapse toward each other, removed the cosmological constant from its equations claiming that was the biggest mistake of his life. Paradoxically, the cosmological constant was recently resuscitated to explain the accelerated expansion of the Universe, it looks like Einstein was right even when he was convinced to be wrong.

4 A Dark Discovery

Meanwhile, astronomers continued to refine the knowledge of our own galaxy and of the external ones. One particular study made by Zwicky in 1938, focused on the velocity of the galaxies belonging to the Coma cluster of galaxies. Assuming some simplifying hypothesis like spherical symmetry and dynamical equilibrium, Zwicky did reach the unexpected conclusion that the mass of the cluster was tens of times larger than computed adding the mass of the individual galaxies. This mass discrepancy did become know as the *missing mass* problem, one of the most outstanding problem in astrophysics, still unsolved.

At the time Zwicky presented his results, it was logic to assume that the missing mass was simply mass that could not be detected by the available optical telescopes, either mass too cold or too hot to emit visible light. It was thought that with sufficiently powerful telescopes sensitive to these wavelengths the mass discrepancy would disappear. Here is worth to remember that some peculiar deviations of Uranus and Mercury from their expected orbit, made astronomers to suggest that still-to-be-discovered planets were affecting their motion. This proved to be the case for Uranus, and Urbain Le Verrier was able to predict the position of Neptune before its discovery at the Berlin Observatory by Johann Gottfried Galle in 1864. What a sensational achievement for science! However, a similar prediction for Mercury—which actually led to the erroneous discovery of Vulcan, a planet supposedly orbiting very close to the Sun—did prove to be wrong. The tiny deviations of Mercury from Newtonian expectations was not due to the existence of another planet. Rather, general relativity, a more refined theory of gravitation was needed to correctly explain its motion.

We recall here that in 1860 Maxwell developed his famous equations, explaining that light is an electromagnetic wave. Being a wave, light needed a medium to propagate, like sound waves that cannot propagate in vacuum. The luminiferous aether, a remarkable weightless and infinitely transparent material filling the whole Universe, was proposed to be responsible for propagating light. A host of experiments were set up to detect it, for instance comparing observations obtained with water-filled and air-filled telescopes! The most famous of all these experiments is the one carried out by Albert Michelson and Edward Morley in 1887, to measure the speed of the earth with respect to aether. This became known as the most famous failed

experiment in the history of physics, the movement of the earth could not be detected. Later, Einstein with his innovative ideas on space and time was able to demonstrate that the experiment failed simply because aether does not exist. It is an unnecessary hypothesis!

In cluster of galaxies we now know that a hot intra-cluster medium exists, as proved by x-ray observations. While this medium largely outweigh galaxies, the missing mass problem was only alleviated, a large discrepancy between observed mass and dynamical mass still remains. Because of this, the missing mass problem was arbitrarily claimed to be solved assuming some invisible and undetectable mass responsible for the discrepancy do exist in clusters. More generally, the dynamical study of all extragalactic objects—spiral and elliptical galaxies, either giant or dwarf, among others—did show a mass discrepancy that is interpreted as due to the presence of unseen mass, baptized *dark matter*.

It is striking to realize that modern astrophysics, in particular modern cosmology, rests on the existence of this unseen dark matter, which should actually outweigh regular matter by a 5 to 1 ratio.

We see here that, like in the case of planets, one part of the problem was solved by discovering the missing mass (the hot intra-cluster medium), though the other part of the problem did not really find a solution. Is the dark matter hypothesis correct, or do we need new physics as for Mercury and aether?

Although there is overwhelming agreement among astrophysicists that dark matter will be discovered sooner or later, at the time of writing it remains an unproven hypothesis and must be treated as such.

5 A Key Discovery that Lead to an Even Darker Discovery

The key discovery from which the modern vision of the Universe was forged, occurred in the sixties. According to the expansion of the Universe discovered by Hubble, in the past galaxies had to be closer together than they are today. Hence, receding sufficiently back in time there must have been a moment in which all galaxies were joined together. This moment was identified with the beginning of the Universe. The modern vision is that all objects were generated in a tremendous explosion, the Big Bang, in which every thing was created (including space and time). Of course, the Big Bang did not generate stars or galaxies. Only a huge amount of energy was released from a state of infinite density and temperature. Expanding at tantalizing speed, as temperature dropped, energy converted into matter, mainly protons, electron and neutrinos. It is estimated that during the first 3 min of existence of the Universe, the temperature and density were sufficiently high that nuclear reaction had time to take place, resulting in the transformation of part of the hydrogen into helium (about 25% in mass), with traces of heavier elements (mostly lithium). As the expansion proceeded, the temperature decreased to reached a point in which protons and electrons could combine to make neutral atoms. This phase occurred about 300 thousands years after the Big Bang. Called “recombination” it correspond to the moment in which

the Universe passed from been filled by an opaque plasma (like the surface of the Sun) to became transparent, allowing photons to travel undisturbed until present day. The oldest signal we might ever hope to receive from any object in the Universe is precisely light emitted at this moment.

The cosmic microwave background (CMB) accidentally discovered in 1964 by Arno Penzias and Robert Wilson, two engineers of the Bell Telephone laboratory assigned to investigate the origin of signals disturbing civil radio transmission, has been interpreted as being the signal coming to us precisely from the moment of recombination [1]. This discovery, worth to them the 1978 Nobel Prize in Physics, is presented as the direct, unquestionable prof of the occurrence of the Big Bang.

The big picture is therefore particularly simple: a big explosion representing the beginning of the Universe, followed by a series of events that culminated in the formation of stars and galaxies, which then passively evolved up to present day [1]. From an astronomical point of view, the Universe is estimated to be very young, with an age of only 13.7 billion years, the sun belonging to the third generation of stars—astronomers count stellar generations backwards, so the very first stars to form after the Big Bang are called generation three, old stars in globular clusters are generation two, and young stars in the disk of the Milky Way including the Sun are generation one. This sequence is amazingly short, it is like if we were living among the nephews of Adams and Eva! By the way, if one wants to retain the concept of creation of the Universe by God, the "Fiat Lux" is obviously identified with the Big Bang.

General consensus were rapidly reached among cosmologist about the correctness of this big picture, and the Big Bang hypothesis has became the undisputed truth, to such an extent that among academics proposing something else is like claiming the earth is flat.

Theoretical considerations suggested that temperature fluctuations at the millikelvin level should be present in the CMB, corresponding to density fluctuations—higher temperatures indicating higher densities. The story goes that the hotter regions, being denser, attract stronger the surrounding material, evolving into more and more massive structures, eventually becoming stars and galaxies. Most important, the exact angular separation on the sky of the fluctuations provides a method to derive the age of the Universe.

Because of this, enormous amount of resources have been spent over the last 4 decades to study the properties of the CMB, culminating with the COBE, WMAP, and Planck characterization of these tiny fluctuations down to an impressive accuracy [2]. A result worth the 2006 Nobel prize for physics to John C. Mather and George F. Smoot (Fig. 1).

This has been recognized as the pinnacle of the Big Bang hypothesis, marking the beginning of the so called "precision cosmology". The level of confidence among modern cosmologists is so high that it is now routine for cosmologists to quote the age of the Universe more accurately than geologists quote the age of the earth.

Is it also routine to forget mentioning that all these results are model dependent resting, for instance, on the unproved validity of the cosmological principle.

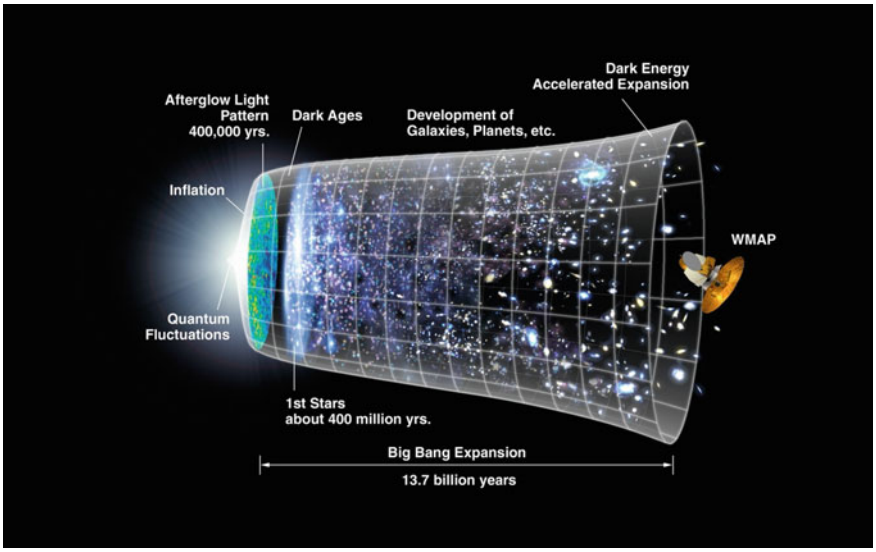

Fig. 1 Sketch representing the evolution of the Universe from Big Bang to present days (credit NASA/WMAP Science Team)

Furthermore, the number of free parameters appearing in the relevant equations is uncomfortably large compared to the number of observables.

It was precisely trying to pin down one of these free parameters that the last major discovery in cosmology was made. Called *deceleration* parameter, it gives a measure of the rate of deceleration of the expansion of the Universe due to the gravitational attraction of the mass within it.

A systematic search involving the Hubble space telescope and all major ground bases facilities was started in the 90s. The aim was to find distant supernovae, standard candles of known luminosity that can be used to measure the geometry of the Universe. The search culminated in 1998 [3] (and references therein) with the discovery that the expansion is actually *accelerating*. Yes, the *deceleration* parameter turned out to be an acceleration.

The problem with an accelerated expansion is that some sort of negative gravity is required to produce it. Of course, no object can produce such an effect, not even the elusive dark matter. The solution was formally found re-introducing a positive cosmological constant in Einstein equations. Constant that is interpreted as due to the existence of a new source of energy permeating the Universe (aether again?) and producing a negative pressure. Baptized "dark energy", this new entity is even less known than dark matter, and unfortunately even more abundant. It should be responsible for nothing less than 70% of the entire energy budget of the Universe.

With this new "epicycle" in place, cosmologists now believe to have completely understood the grand design of the Universe, only details remain to be defined. It is difficult to refrain making a parallelism with the situation in which physics were in

1900, resumed by the famous sentence pronounced that year by Lord Kelvin: *"There is nothing new to be discovered in physics now. All that remains is more and more precise measurement."*

We all know that few years later physics was completely revolutionized by quantum mechanics and relativity. Should we expect the same for cosmology?

Meanwhile, Saul Perlmutter, Brian P. Schmidt, and Adam G. Riess, **t**hree leading member of the supernovae search team, shared the 2011 Nobel prize for physics.

6 Multi Messenger Astronomy

One basic ingredient of general relativity is the notion of gravitational waves, responsible for transmitting the gravitational field, similarly to electromagnetic waves that transmit the electric and magnetic field. At the time Einstein introduced this concept, no one could prove or disprove it. A full century of amazing technological advancements had to be waited. Confirmation arrived on 14 September 2015, when the LIGO project announced the very first detection of gravitational waves, soon followed by the independent confirmation from the Virgo project on 11 February 2016. The received signal was named GW150914. The waveform detected, reflects exactly the predictions of general relativity for a gravitational wave produced by a spiral trajectory and final merging of two black holes with mass equal to 36 and 29 solar masses, and the resulting "ringdown" of the resulting single black hole [4] (Fig. 2).

It was also the first observation of the merger of a binary system of black holes, thus demonstrating the existence of stellar-mass black hole systems and that such mergers can occur within the current age of the universe.

A following detection event GW170817 in August 2017 [5] gave rise to one of the largest observation campaigns ever carried out by the international astrophysics community, using every possible telescope from the ground and from space. Scientists around the world have searched for the light emitted during this event at all wavelengths, from radio to gamma rays.

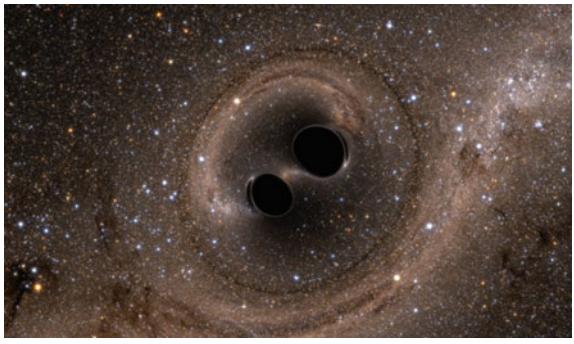

Fig. 2 LIGO computer simulation image showing the merger of two black holes (credit Simulating eXtreme Spacetime project)

It was a success because, only two seconds after the detection of the gravitational signal, the Fermi and INTEGRAL satellites discovered a weak gamma-ray burst coming from the same sky region. The subsequent and intensive search from ground-based telescopes all around the world identified a bright new source in the galaxy NGC 4993, coinciding with the position and distance indicated by the gravitational wave signal. Once the position of the source in the sky was identified with extreme precision, the subsequent monitoring at all frequencies made it possible to follow the details of the complex explosive phenomenon and track it down to be generated by the merger of two neutron stars. Rainer Weiss, Barry Barish, and Kip Thorne were awarded the 2017 Nobel prize for physics for the observation of the gravitational waves that confirmed a prediction from Albert Einstein theory of general relativity.

During the same year, the IceCube experiment in Antarctica detected a high energy neutrino EHE 170922A, of extragalactic origin (IceCube collaboration 2018). The neutrino emission could be associated to the bright radio and γ-ray emitting BL Lac object TXS 0506 + 065, located at a redshift $z = 0.336$ [6], as measured by the 10.4 m Gran Telescopio Canarias, at present the world largest optical telescope.

Perhaps the most important aspect of gravitational waves and neutrinos detection is that they represent the first signals other than electro-magnetic radiations that we have ever received from the heavens. The multi-messenger astronomy was born: the capacity of the modern science to study the Universe using more than a carrier of information beside light. These new tools open new windows that will help to deepen our knowledge of both astrophysics and fundamental physics (Fig. 3).

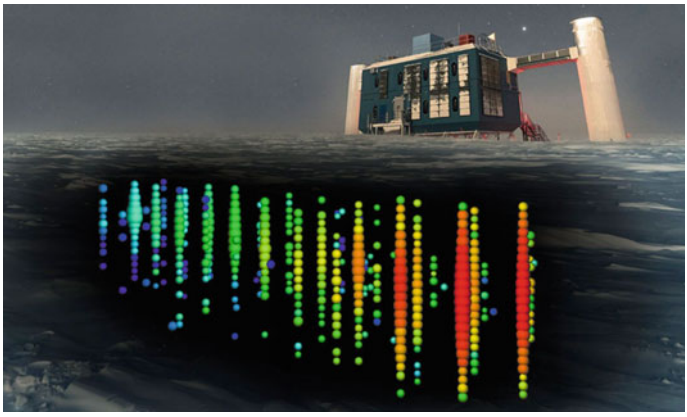

Fig. 3 Artistic representation of IceCube detection of a high energy neutrino (Credit: IceCube Collaboration)

7 Modern Astronomy

Invisible by definition, black holes, another prediction of Einstein theories, are exotic astronomical objects that can be detected investigating their gravitational effects on nearby objects, or when accreting material. They are thought to be the ultimate engine powering quasars, the most powerful sources in the Universe.

Starting from 1992, Reinhard Genzel and Andrea Ghez made use of the high spatial resolution performances provided by adaptive-optic instruments installed in the 10 m class telescopes Keck and VLT, to study the galactic center. In a very extended monitoring campaign lasted for more than two decades [7], they demonstrated that an invisible and extremely heavy object governs the orbit of stars at the center of our galaxy: a quiescent supermassive black hole. For this discovery they were awarded the 2020 Nobel prize for physics.

The Event Horizon Telescope (EHT) is an international project established in 2006 with the aim of studying the supermassive black hole located at the center of nearby galaxies. To this end, a system consisting of several radio telescopes connected by means of the Very Long Baseline Interferometry (VLBI) technique was used. In VLBI, a signal emitted by a radio source is collected by several radio telescopes located on Earth. Each antenna receives the data with a certain time delay that depends on its exact position in the array. Data are then synchronized with the aids of atomic clocks and combined, to reconstruct the information as if provided by a single dish antenna, with spatial resolution set by the maximum distance of the individual telescopes.

After the first experiments in 2006 the EHT network has expanded to include almost all the new submillimiter observatories that have entered into operation in the meantime, including the more powerful Atacama Large Millimeter Array (ALMA). Inaugurated in 2013, ALMA is a radio telescope composed by an array of 66 Antennas operating in the millimeter and sub-millimeter wavelength range, from 350 micron to 10 mm. It represents the natural bridge between classical radio astronomy and optical-infrared astronomy. The advent of ALMA opened a completely new window on the investigation of all the physical phenomena happening in the cold matter or hidden in dust and gas clouds as the formation of planets, proto-stars and proto-galaxies. ALMA on its own, with the possibility to arrange its antennas over an area of 16 km in diameter in the Chajnantor plateau, provides the incredible resolution of 10 milliarcsecs (5 times better than the Hubble Space Telescope) and with a collecting area of 4700 square meters it is the most powerful sub-millimeter telescope currently in operation (Fig. 4).

The radio telescopes making up the EHT network are thousands of kilometers from each other, giving an angular resolution of approximately 20 micro-arcsec, capable of resolving the event horizon of the aforementioned black holes. The first direct VLBI image of a black hole, the one at the center of the giant elliptical galaxy M87 in the Virgo cluster, was released on April 11, 2019 [8] (Fig. 5).

For the relevance this kind of investigations do have, they project straight into the future of astronomy, into the epoch of the 40 m class optical telescope.

Fig. 4 Aerial view of ALMA taken from a small remote-controlled hexacopter (credit EFE/Ariel Marinkovic)

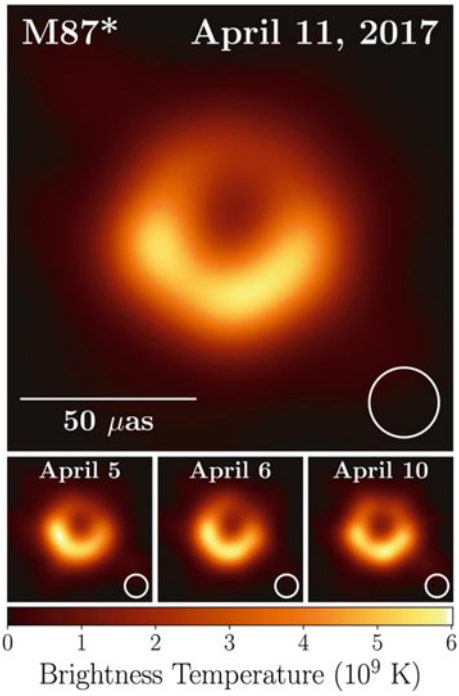

Fig. 5 The first image of the black hole at the center of M87 galaxy (credit Event Horizon Telescope)

Fortunately, a lot of matter in the Universe is visible. There are billions of observable galaxies, and each galaxy contains billions of stars. It is reasonable to imagine the Sun can't be the only star hosting a planetary system, rather, they should be quite common. This goes together with one of the most pressing questions humans have ever had: are we alone in the Universe? While to provide a definitive answer to this question is still not possible, great steps forward have been made thanks to the tremendous technological innovations of the last few decades.

The first detection of a planet orbiting a star other than the sun, an exoplanet, occurred in 1992 when the astrophysicists Aleksander Wolszczan and Dale Frail discovered three exoplanets orbiting the pulsar PSR1257 + 12 [9], certainly an unexpected environment for planets.

This was just the beginning. Thanks to the methodology developed by the Swiss astronomer Michel Mayor in the 90's, in 1995, himself and Didier Queloz detected the first exoplanet around a "normal" (main sequence) star, 51 Pegasi [10]. The planet, named 51 Pegasi b, has approximately half the mass of Jupiter and whizzes around its parent star in just over four Earth days, lying almost eight times closer to it than Mercury is to the Sun. In 2019 Michel Mayor and Didier Queloz were awarded the Nobel prize for physics for their discoveries.

Since 1995, this area of astronomy has become one of the most dynamics of all. Basically, every star that could be studied with sufficient details is found to have planets, definitely more than initially expected.

Exoplanets are discovered using indirect methods, either exploiting the gravitational effects they produce on their host star, or observing the tiny eclipse produced when the planets transit over the disk of its host star. Because of this, radial velocity measurements have reached the cm/s accuracy in order to detect the minute wobble a planet produces on the star radial velocity curve, and one-part-per-million photometric precision is now routinely achieved to detect the exoplanet transit. This is particularly important because observing the eclipse in different bands, or spectroscopically, puts constraints on the chemical composition of the exoplanet's atmosphere.

From the first detection, and thanks to dedicated projects like the Kepler space mission, at the time of writing over 5000 exoplanets around 3600 planetary system have been discovered.

Several of these worlds are located in the so called habitable zone. For instance, Kepler-1649c orbiting a star at 300 light years from the earth, four of the seven terrestrial planets discovered around the star Trappist-1 at 40 light years from us, and Próxima Centauri b at only 4.2 light years. At present, more than 30 planets have been confirmed to orbit in the zone in which they can support liquid water in their surface, given sufficient atmospheric pressure.

It is estimated that in the Milky Way at least 17 billion Earth-like exoplanets exist, of which more than 300 millions might be habitable planets.

The hardest way to detect an exoplanet is to try to see it directly, because of the extreme contrast between the bright star and the faint planet. To directly detect the planet in the optical, the only method is to physically block the starlight (using some sort of coronagraph), masking the star while allowing nearby planets to shine

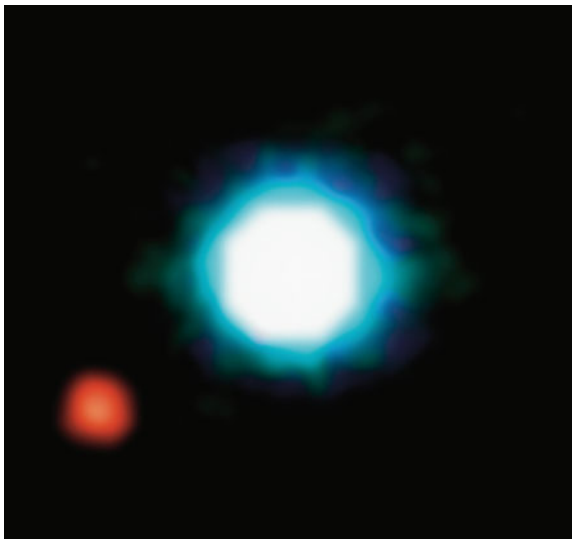

Fig. 6 First direct image of an exoplanet (red) obtained with NACO@VLT (credit ESO)

through. It is furthermore necessary to remove the turbulence introduced by the earth's atmosphere, which smear out the image of the star preventing the direct observation of the planet. This became possible with the recent advent of 10 m class telescopes equipped with Adaptive Optics, instruments able to correct the effects of the atmosphere in real time. The first direct image in the near infrared of an exoplanet has been obtained in 2004 using NACO at the ESO-VLT Paranal observatory [11] (Fig. 6).

Not surprisingly, exoplanet imaging is one of the main drivers for building the European Extremely Large 39 m telescope, planned to enter operation in 2025.

The existence of extra-solar planetary and proto-planetary systems has been independently confirmed by the most recent discoveries of the Atacama Large Millimeter Array. The first detailed image of a planetary disk, obtained in 2014, was observed around HL Tau2, a million-year-old Sun-like star located approximately 450 light-years from earth. After that, several other extra-solar systems have been revealed. The quality of ALMA data is so high that it was recently possible to unequivocally detect the presence of a disk around an *exoplanet* [12], which is supposed to be forming a moon!

Finally, the recent discoveries by ALMA of complex organic carbon molecules in protoplanetary systems around stars and in the Magellanic clouds, it is easy to imagine that the search for places suitable for the development of life will be one of the main field of research in the next decades.

8 High-Technology Challenges

The tremendous technological acceleration we witnessed in the last 30 years enormously increased the capacities of the astronomical investigation. This has produced an epochal leap in the quality of the results produced, while opening a plethora of new questions about the real physical nature and evolution of the Universe.

From the technological point of view, the main drive for this great revolution were fundamentally three: the ability to produce large optical surfaces enormously increasing the collecting area for photons, the ability to compensate the effects of turbulence introduced by the atmosphere, and, last but not least, the creation of large international consortium.

Let's start with the mirrors. For many decades the large 5 m telescope at Mount Palomar represented the limit in the construction of large telescopes. For this telescope, as for all those built before the 1980s, optical stability was obtained through very rigid and heavy structures that guaranteed the minimum deformations of the optical train and therefore of optical aberrations, in all orientation of the telescope. This obviously included a monolithic structure for the primary mirror that in the case of the Palomar telescope weights 20 tons and is 60 cm thick.

The same construction concepts guided the construction of all telescopes of the 4-m class. In the 1980s the development of electronic detectors (CCDs) offered great possibilities for progress in astronomical observation without the need to build larger telescopes. In practice, with CCDs we go from the 1% quantum efficiency of photographic plates, to 90 plus percent. And the real increase in efficiency is even larger than this, because photographic plates suffer of reciprocity failure, meaning that doubling the exposure time you get less than twice the signal. In the case of the very long exposures typical of astronomical observations, the quantum efficiency approaches zero. All this was solved by the introduction of CCDs. To give an example, the tremendous observational effort made by Hubble to discover the expansion of the Universe, carried out at the then largest telescope in operation, the Mount Wilson 2.5 m telescope, could be done easily today with an amatorial 25 cm telescope equipped with a good CCD. For this reason, the conceptual development of new telescopes that could go beyond the structural limits of the Palomar one remained dormant for decades.

Only when the efficiency of the detectors could not be further improved, did the problems associated with the need to increase the size of the telescopes arise again in its entirety. Simply scaling the telescope structure to a larger size was not practicable mainly due to the negative impact that large mass structures produced on the observing conditions. The Russian BTA-6 telescope, with its 6 m monolithic primary mirror, is a good example of this. The BTA-6 telescope was, however, interesting for being the first large telescope with an altazimuthal mount, which is now the standard for all large instruments.

Other conditions been equal, for instance the level of light pollution affecting the observations, the effectiveness of a telescope is basically measured by two parameters: the collecting area (to detect faint targets) and the ability to concentrate light (spatial resolution).

In the focal plane of an ideal telescope free from aberrations, the light of an astronomical point source is not a point, rather it is spread on a disk (called Airy disk). The diameter of this disk represent the minimum angular separation to distinguish 2 nearby sources, for example to resolve a double star. The larger the mirror, the smaller the disk and higher the resolution. A 10 cm telescope in the optical band has a resolution of about 1 arcsec, and one of 5 m should give us a resolution of 0.02 arc seconds. Unfortunately, this does not happen for a reason astronomers call "seeing".

Seeing is the combined effect of atmospheric turbulence (air moving above the telescope) and more local turbulence generated by the dome and the telescope's structure. Thus, the theoretical diffraction limit is degraded to the seeing figure which can be of several arcsec. Indeed, till the 1980s it was believed that the natural seeing figure (moving atmosphere) could not be less than 1 second of arc. In other words, a large telescope could not have, and in practice didn't have, a better spatial resolution than a 10 cm one (though it would be able to detect much fainter sources). Precisely to avoid this limitation, the Hubble Space Telescope (HST) was built and put into orbit. In the meantime, for large terrestrial telescopes, the effects of local turbulence were limited (with little success) by building the telescopes on structures elevated from the ground as much as possible and insulating the huge domes to avoid the accumulation of heat. The 4 m Mayall telescope at kitt peak observatory, with its 60 cm thick primary mirror and an impressive dome, or the ESO's 3.6 m at La Silla observatory are vivid example of the technical astronomical wisdom of the sixties and seventies.

9 Better Sites and New Technology Telescopes

In the 1980s new astronomical sites such as Mauna Kea in Hawaii at a height of 4200 m, were identified. Far from light pollution, these sites were selected for having atmospheric seeing less than one arcsecond, often reaching 0.5 arcsec or better. To take advantage of this exceptional sites, it was therefore necessary to reduce the damage to the image quality made by the dome and the telescope structure, while maintaining or increasing the diameter of the primary mirror. Thus, engineers moved from the concept of a rigid heavy mirror to that of a thin and deformable one, capable to compensate the deformations induced by the weight orientation and varying temperature in real time, by continuously changing the shape of the primary mirror. The *active optics* was born. The first example of a new generation telescope was the ESO's 3.5 m NTT installed at Silla. Conceptually developed by Ray Wilson, the NTT is made up of a lightweight primary mirror (only 10 cm thick), deformable through actuators that push and pull on its back. The equatorial mount was also abandoned in favor of the altazimuthal one. This is much more compact for having

only an horizontal and a vertical axis of rotation, instead of the inclined axis (parallel to the earth axis) characteristic of the equatorial mount. This, combined with shorter focal length of the primary mirror (passing from typical apertures f/D > 3 to f/D < 2), resulted in extremely compact telescope and consequently smaller dome buildings. All these solutions, together with active extraction of the heat generated by motors and electronics, made it possible to obtain images with spatial resolution as good as 0.3 s of arc, demonstrating all the effectiveness of the new technologies adopted.

The ability to use actively controlled thin and deformable mirrors has become a constant. All the subsequent 8 m class telescopes, like the ESO VLT, or the University of Arizona twin 8.4 m mirrors of Mt. Graham's Large Binocular Telescope are constructed according to this. However, the 8 m monolithic mirrors will possibly remain the largest ever made. The constant increase in computer power allowed for faster and more accurate control of active optics, opening the way to another important change in the telescope design. The two Keck 10 m telescopes located at Mauna Kea in Hawaii, for the first time demonstrated that it was feasible to build a large primary mirror joining together several smaller mirrors, maintained in position to the required nano-meter accuracy by an active control. Being segmented, the diameter and the thickness of the primary mirror are no longer correlated, allowing for the construction of extremely thin mirrors. For instance the primary mirror of the Spanish GTC 10.4 m telescope has segments only 8 cm thick. All large telescopes under development do foresee the use of segmented primary mirrors. The ESO's E-ELT 39 m giant telescope, under construction at Cerro Armazones in northern Chile, will boast a primary mirror made of almost 800 segments (Fig. 7).

Fig. 7 The segmented primary mirror of the 10.4 m GTC telescope, at present the largest optical-infrared telescope in operation. Note the different reflectivity of the segments due to oxidation, each segment is re-coated once a year (credit G. Gomez Velarde)

Beyond the well demonstrated technical feasibility, a fundamental point in favor of large ground-based telescopes is their competitiveness with space telescopes, which due to the payload constraints for putting them into orbit have necessarily relatively small mirrors. Space telescopes do have, however, the enormous advantage of operating in the absence of atmospheric distortions, which represents an incredible advantage from the point of view of resolution. Since the depth reachable by a telescope depends on both the collector area and how much the light is concentrated on the focal plane, ground based telescopes cannot be satisfied with the resolutions of a few tenths of an arcsec, as permitted by the use of active optics. The keystone is therefore the possibility of correcting in real time the distortions introduced by the atmosphere on the images.

10 Adaptive Optics

Already in the 1950s it was assumed that the effects of seeing could be reduced, by introducing a deformable optical element (adaptive optics) that can compensate in real time for the distortions induced by the atmosphere. The concept itself is simple and is based on the analysis of the shape of a bright star in the field of view, the distortions of this are translated into a correction matrix that actuators apply to a deformable mirror, introducing a deformation opposite to the one produced by the atmosphere. In this way smoothing out the wavefront distortions. If it was so conceptually simple, why did it take half a century to see it applied to astronomical observation? The explanation lies in the extreme rapidity of the evolution of atmospheric turbulence. Corrections have to be computed and applied within a few milliseconds, typically with a 50 Hz frequency, which requires a remarkable computing power and highly flexible optical elements that can support the stress induced by the continuous deformation. Technology from the 90s onwards has provided us with these last two ‘capabilities’ and adaptive optics is now a reality.

However, this revolutionary technique remained limited by the needs of a bright reference objects (a stars) in the field of view of the instrument, a limitation overcome creating “artificial stars”. In practice, one or more powerful laser beams aligned with the telescope are used to excite the sodium layer present in the ionosphere at an height of about 90 km, effectively creating bright spots that can be used as artificial stars. The analysis of the wave front coming from these artificial stars allows to correct the effect of the atmosphere in any point of the sky (Fig. 8).

The die was cast, large ground-based telescopes with adaptive optics technology could compete with space-based telescopes and produce deeper and more detailed observations than the latter thanks to the larger size of the primary mirror.

The measurements of the stellar orbits in the high-density regions around the center of our galaxy, or the first image of an exoplanet we talked about earlier, were obtained thanks to this technology. It is therefore not surprising that all new large terrestrial telescope designs will make extensive use of adaptive optics. This technology initially limited to dedicated adaptive optics modules, deformable mirrors

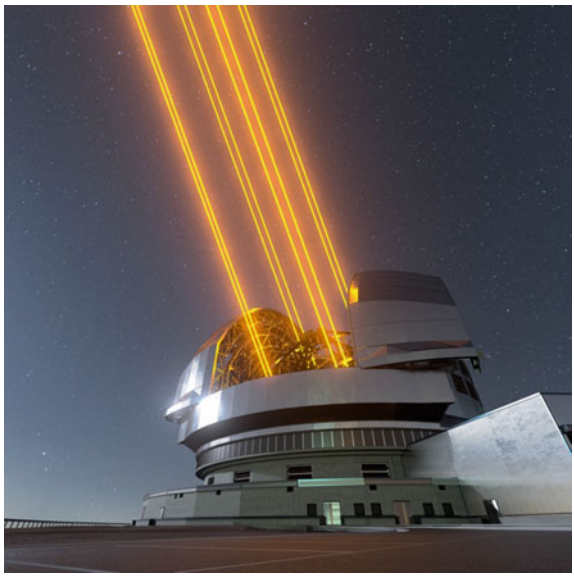

Fig. 8 ELT artistic view of the shooting laser guide star system (credit ESO)

of a few centimeters in diameter placed within individual instruments, will become an integral part of the main optical system. Yepun, one of the fourth telescopes of the VLT, has been already equipped with an adaptive M2 secondary mirror of 1.1 m in diameter. In the ELT of ESO this will go much further, the adaptive module being made by the M4 and M5 mirrors having 2.4 and 2.2 m of diameter, respectively. M4, a true technological marvel, is the largest deformable mirror ever built. Only 2 mm thick, it will be a levitating surface maintained at a distance of about 1 tenth of a millimeter from the reference surface. Its shape will be checked 70 thousand times per second with an accuracy of 0.00005 mm. Five thousand coil actuators will deform it 1000 times per second, making the surface vibrate like if it were a loud-speaker, allowing for the compensation of the rapid atmospheric turbulence. M4 will be coupled with M5 which ensures image stabilization by compensating the perturbations caused by the telescope mechanics and wind vibrations through a rapid tip and tilt system. This jewel of optomechanical engineering will cancel out the atmospheric turbulence almost completely, allowing the E-ELT to reach its diffraction limit of a few milliarcsec, transforming a terrestrial telescopes into virtual space telescopes (Fig. 9).

11 Operational Challenges

Hand in hand with the technological revolution, a real structural and in some ways industrial revolution of astronomical investigation has been generated.

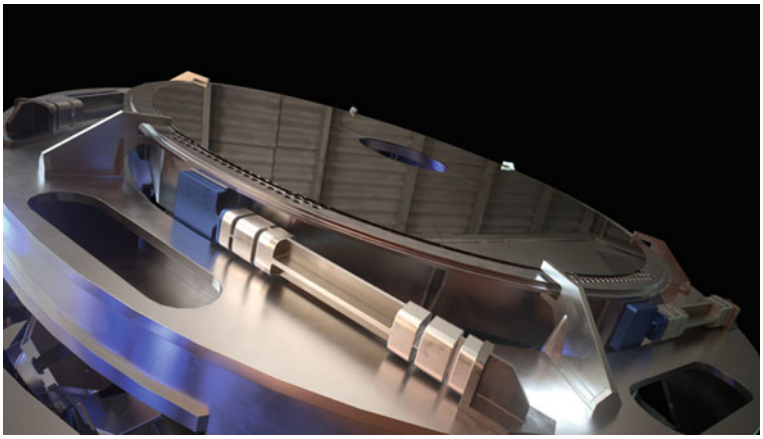

Fig. 9 ELT thin deformable M4 Adaptive Optic mirror (credit: ESO)

The main reason is to be found in the need to maximize the efficiency and productivity of the observing time at large telescopes. These concepts, which have always been pillars of space missions, only became part of ground-based astronomy with the advent of large 10 m class telescopes.

These are expressed at the operational level in four fundamental actions: careful selection of the proposed projects to extract the most of the available instruments, observations carried out by expert resident astronomers, creation of easily accessible data archives, and finally optimization of the data reduction process (pipelines). The last two represented a real revolution for observational astronomy. The possibility of using the acquired data for scientific purposes other than the original one, together with the possibility of using data from different telescopes and instruments has enormously increased the effectiveness of the astronomical investigation. At present, the large amounts of data collected by modern astronomical observatories are cataloged and archived in large databases accessible online, which in many cases offer both the raw and reduced data, produced by dedicated pipelines. The virtual observatory (www.virtualobservatory.org) or the HST archive (archive.stsci.edu) are good examples of this.

However, the traditional way to extract information from the observed data no longer works. Even though what in the past required dedicated workstations can now be done with a smartphone, the amount of data has increase beyond belief. The data flow of ALMA reaches 1.5 terabyte per second! A dedicated supercomputer, the Atacama Compact Array Correlator System, was developed to handle this data rate, process the information in real time, and send it to the archive. The system is capable of performing 120 trillions of operations per second. Archives and pipelines are the natural complement to the incredible quality of the data modern astronomical instruments produce. The bottom line of all this, is an enormous increase of the discovery rate. A result further facilitated by the presence of expert resident astronomers operating modern telescopes in service mode. This optimizes the use of observing time,

making possible for instance to carry out program requiring fast reaction time. The case of γ-ray burst (GRB) that can appear at any time in any position of the sky is particularly instructive. GRB are initially discovered by γ-ray satellite, which in real time pass the approximate position of the γ-ray source to x-ray satellites. These immediately point the region of interest and identify an x-ray transient determining its coordinates accurately enough for an optical telescope to put it on the slit of a spectrograph and measure its distance. All this within minutes from the initial discovery. Obviously, only a team of resident astronomer can carry out this observations. Fully robotic telescopes are also another reality that allows for systematic monitoring of specific targets.

Telescope maintenance must also be optimized in all modern observatory to maximize efficiency. The typical example here is the need of recoating the mirrors to cope with atmospheric oxidation of their reflecting surface.

At present this could be considered a normal maintenance activity, carried out typically once a year. However, it will definitely become a challenge for the 40 m telescopes entering operation in the next decade. At the ESO's ELT, several segments will have to be recoated every day in order to maintain the requested performances of its primary mirror consisting of 798 segments. An industrial style "assembling line" will have to be put in place to cope with this.

Clearly these huge steps forward were made possible not optimizing here and there but putting together all the ingredients necessary to perform astronomy at the highest level, creating a synergy that did miss for too long in this field of science. One of the last example of a global effort is the development of the telescope that has to replace the HST.

12 What Next?

By the time this text will appear in print, the joint NASA, ESA and CSA James Webb Space Telescope, by far the most expensive telescope ever built, will have reached L2, the Lagrangian point 2 located about 1.6 million kilometers away from earth. There, it will join the ESA GAIA astrometric satellite and the Russian-German high-energy astrophysics observatory Spektr-RG. Other famous satellites have accomplished their mission at L2, like NASA's WMAP and ESA's Herschel Space observatory. This stable point (with respect to the earth) is used because at that distance the background produced by scattered light in the high atmosphere is completely absent. Surprisingly enough, even though it is orbiting at ~ 530 km of altitude, the HST is still affected by scattered light from the atmosphere. At L2, the sky background is about 100 times fainter than seen by HST.

Because of this and thanks to its large 6-m primary mirror, optimized for optical infrared observations, the JWST is expected to revolutionize astronomy, possibly even more than HST did. Of course, there are many more fully fledged telescopes out there in space, all the x-ray satellites for instance, and some of them are already at L2. But historically, optical telescoped always had an additional emotional appeal that

Fig. 10 James Webb Space Telescope during assembly (credit NASA)

make them particularly attractive. This mostly because they can deliver unparalleled images of exquisite quality of "famous" objects, and the comparison with ground-based images of the same targets make it clear why it was worth to spend all the money these telescopes cost.

The JWST is meant to give important contributions to all fields of astronomy, from solar system targets and exoplanet orbiting nearby stars, all the way to the most distant sources known.

Indeed, it is expected to discover the very first generation of stars and galaxies ever appeared in the Universe, marking the end of the so called "dark age" at $z > 10$ (Fig. 10).

Here astronomers are moving on uncharted territory, never studied before. Thus, it is possible that the JWST might be able to observe even further than the first galaxies, demonstrating that indeed before that there was nothing. This would be simply sensational!

13 Conclusions

Expectations are high, but of course, when science moves ahead into new territories, surprises are always around the corner and, to be honest, it is difficult to make predictions of what the new generation of telescopes will discover. Will the expectation of modern cosmology and the Big Bang hypothesis be confirmed? If so, this would automatically imply that the elusive dark matter and dark energy do really exist. Many astrophysicists agree that this is going to be the case. But, for instance,

Fig. 11 ESO's Paranal Observatory VLT photo @ Mike Struik and Paola Catapano

HST data confirm that fully formed galaxies and super-massive black holes powering quasars are already in place at $z = 6$, when the universe was only 1 billion years old. It is difficult to explain how these objects could have formed in such a short period of time. This already poses big problems for the standard model. Will the JSWT and ELT astonish the scientific community finding that even beyond $z = 10$ there are fully formed galaxies and quasars? If so, a major revision of our vision of the Universe will be necessary.

History has taught us that new technological developments always come with both answers to known problems and questions associated to new discoveries.

As scientists, we hope the near future will bring many more question marks than ending periods (Fig. 11).

References

1. A.A. Penzias, R.W.Wilson, A measurement of excess antenna temperature at 4080 Mc/s. ApJ **142**, 419–421 (1965). DOI: 1086/148307
2. G.F. Smoot, C.L.Bennett,, A. Kogut et al., Structure in the COBE differential microwave radiometer first-year maps. Astrophys. J. Lett. **396**, L.1 (1992). https://doi.org/10.1086/186504
3. B.P. Schmidt, N.B. Suntzeff, M.M. Phillips et al., The High-Z supernova search: measuring cosmic deceleration and global curvature of the universe using type Ia Supernovae. Astrophys.

J. **507**, 46–63 (1998). https://doi.org/10.1086/306308
4. LIGO Scientific Collaboration and Virgo Collaboration. B.P. Abbott, R. Abbott, T.D. Abbott et al., GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. Phys. Rev. D **93**, 122003—Published 7 June 2016 https://doi.org/10.1103/PhysRevD.93.122003
5. LIGO Scientific Collaboration and the Virgo Collaboration: B.P. Abbott, R. Abbott, T.D. Abbott et al., On the progenitor of binary neutron star merger GW170817. ApJL **L40**, 850 (2017) https://doi.org/10.3847/2041-8213/aa93fc arXiv:1710.05838v2
6. S. Paiano, R. Falomo, A. Treves et al., The redshift of the BL Lac oObject TXS 0506+056. ApJL **854**, L32 (2018). https://doi.org/10.3847/2041-8213/aaad5e
7. S. Gillessen, F. Eisenhauer, S. Trippe et al., Monitoring stellar orbits around the massive black hole in the galactic center. ApJ **692**(2), 1075–1109 (2009). https://doi.org/10.1088/0004-637X/692/2/1075
8. The EHT Collaboration et al.,First M87 event horizon telescope results.VI. The shadow and mass of the central black hole. ApJ **875**, L6, 1–44 (2019). Open Access https://doi.org/10.3847/2041-8213/ab1141 arXiv:1906.11243v1
9. A. Wolszczan, D.A. Frail, A planetary system around the millisecond pulsar PSR1257 + 12. Nature **355**, 145–147 (1992)
10. M. Mayor, D. Queloz, A Jupiter-mass companion to a solar-type star. Nature **378**, 355–359 (1995). https://doi.org/10.1038/378355a0
11. G. Chauvin, A.M. Lagrange, C. Dumas et al., A giant planet candidate near a young brown dwarf. A&AL **425**(2), L29–L32 (2004). https://doi.org/10.1051/0004-6361:200400056
12. M. Benisty, J. Bae, S. Facchini et al., A circumplanetary disk around PDS70C. ApJL **916**(1), idL2 (2021). http://doi.org/https://doi.org/10.3847/2041-8213/ac0f83. arXiv:2108.07123

Other Worlds in the Cosmos: From Philosophy to Scientific Reality

Michel Mayor, Emeline Bolmont, Vincent Bourrier, David Ehrenreich, and Christoph Mordasini

Do there exist many worlds, or is there but a single world? This is one of the most noble and exalted questions in the study of Nature.
Albertus Magnus (circa 1200–1280)

1 Other Worlds in the Cosmos—An Ancient Philosophical Questioning

It is amazing to consider that the question of the plurality of Worlds in the universe was already discussed in the Antiquity by Greek philosophers. In a very famous letter of Epicurus (341–270 BC) we can read, "Worlds are in an infinite number some of them similar to our own one, some others being different…living species, plants and all the other visible things could exist in some worlds and could not in others."

But, although the original texts have disappeared, this question was already debated by the Ionian philosophers as early as the sixth century BC. It is also interesting to consider the contribution of the Atomists Leucippus and Democritus in the

M. Mayor (✉) · E. Bolmont · V. Bourrier · D. Ehrenreich
Département d'Astronomie, Université de Genève, Ch. Pegasi 51b, 1290 Versoix, Switzerland
e-mail: michel.mayor@unige.ch

E. Bolmont
e-mail: emeline.bolmont@unige.ch

V. Bourrier
e-mail: vincent.bourrier@unige.ch

D. Ehrenreich
e-mail: david.ehrenreich@unige.ch

C. Mordasini
Space Research & Planetary Sciences Division, Physikalisches Institut, University of Bern, Gesellschaftsstrasse 6, 3012 Bern, Switzerland
e-mail: christoph.mordasini@unibe.ch

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_17

Fig. 1 In the Doppelmayr Atlas of 1742, some distant stars, beyond the solar system are surrounded by planetary systems, (see in the figure on the right, the enlargement of a fragment of the Atlas page)

following century. If matter is made up of atoms, it is likely that nature has been able to agglomerate them in such a way to constitute an infinite number of worlds.

The interested reader will be able to appreciate the complex and continuous evolution of the ideas in the monography of Connes [1].

The question of the plurality of worlds in the universe has been continuously present during the last two millennia. We can, for example, quote this sentence by the philosopher and theologian Albertus Magnus (circa 1200–1280), "*Do there exist many worlds, or is there but a single world?*". This is one of the most noble and exalted questions in the study of Nature.

More recently the two major contributions of Emmanuel Kant (1755) in his Universal Natural History and Theory of Heaven and Pierre-Simon Laplace in his Exposé du Système du Monde must be mentioned. Both contributions introduce the notion of protoplanetary nebula, having noticed that all planets are moving in the same plane and sense of rotation (Fig. 1).

1.1 Change of Paradigm During the Second Half of the Twentieth Century

How many planetary systems are there in the Milky Way, hosted by the 200 billons of stars? How many planets are like our Earth? It is interesting to look at the astronomical literature of the twentieth century for estimations of planetary systems in the Milky Way. Before 1943, the estimations were between zero and at most a few. It was supposed that the formation of the protoplanetary nebulae results from the close encounter of two stars. The very low probability of such an event (close to zero!) is at the origin of these pessimistic estimates. In the early 1940s, claims of planet discoveries around some of the closest stars to the Solar System (claims which were later found to be erroneous) produced a complete paradigm shift, with estimates

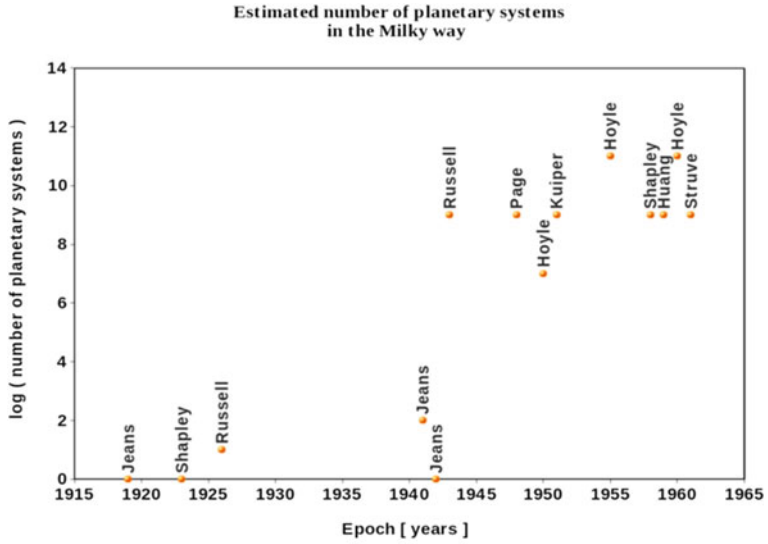

Fig. 2 Estimated number of planetary systems in the Milky Way (adapted from [2]). An illustration of the change of paradigm in the middle of the twentieth century

of planetary systems in our Galaxy as large as hundreds of billions, (Fig. 2). It is interesting to note that this shift in paradigm was the result of spurious detections of planetary systems!

However, in the early fifties, Otto Struve gives the correct explanation for the origin of these planetary nebulae, "*… the lack of rapid axial rotation of normal solar-type stars … suggests that these stars have converted their angular momentum from axial rotation to angular momentum from the orbital motion of the planets. Therefore, there can be many planet-like objects in the galaxy*" [3].

Young stars formed by gravitational collapse of turbulent giant molecular clouds should have extremely large rotational velocities. However, the observed rotational velocities of stars at the bottom of the main sequence (stellar masses less than about 1.2 times the solar mass) are extremely small. Otto Struve suggests that the excess of angular momentum, if not present in the stars themselves, should be present in the protoplanetary nebulae. Consequently, protoplanetary disks are byproducts of the stellar formation itself and we can anticipate that most of stars (if not all of them) should host planetary systems. In the 1970s, an excess of infrared luminosity in the spectra of very young stars finally revealed the presence of protoplanetary disks (Fig. 3).

"The typical disk has an angular momentum comparable to that generally accepted for the early solar nebula, but very little stored energy, almost five orders of magnitude smaller than that of the central star. Our results demonstrate that disks

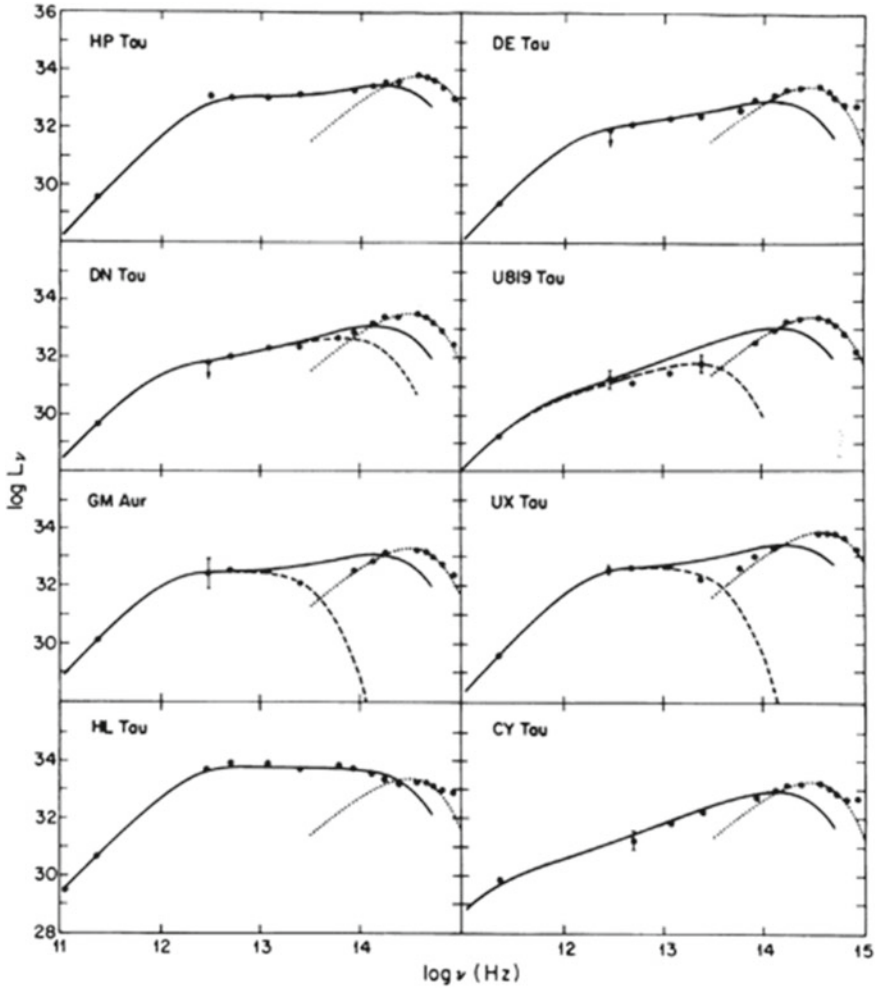

Fig. 3 Spectral energy distributions of eight very young stars. The excess of near-infrared dust luminosity at lower frequency than the stellar's typical own emissions is important (Credit [4])

more massive than the minimum mass of the proto-solar system commonly accompany the birth of solar-mass stars and suggest that planetary systems are common in the Galaxy" [4].

Then in 1995, direct imaging of very young stars moving out of the Orion Nebula (Fig. 4) showed that most of them are surrounded by disks of dust and gas [5].

No doubt at all, most (if not all) stars should host planetary systems.

During the past three decades, improvements in astronomical instrumentation and the development of new observational techniques made it possible to transform the old philosophical concept of "plurality of worlds" in the universe into an active field

Fig. 4 Images of protoplanetary disks: The Hubble Space Telescope reveals protoplanetary disks around very young stars of the Orion Nebula (Credit [5])

of modern astrophysics. Today, more than 4000 exoplanetary systems have been detected and we have even been discovering planets in the so-called habitable zones of host stars.

2 Doppler Spectroscopy as a Path to the Detection of Exoplanets

The relative ratio between the stellar and reflected planet luminosity is so large (1 billion in the visible) than the direct imagery of planetary systems was not possible since the second part of the last century.

Early astrometric research sought to detect the variation in position of stars around the center of gravity of planetary systems. Such an angular displacement is extremely small because of the contrasted mass ratio between stars and planets. Since the middle the twentieth century, all the claims of astrometric detections of exoplanets have been shown to be erroneous (planet hosted by 61 Cygni, [6],hosted by Lalande 21,185, [7], hosted by the Barnard's star, [8] and more recently a planet hosted by VB10, [9]). That situation is expected to drastically change from 2022 with the ESA astrometric mission GAIA [10].

The possibility of detecting the gravitational influence of orbiting planets on the radial velocities of stars was suggested long before the Doppler technique was precise enough to allow such measurements [3, 11]. In the eighties, several teams explored the possibility to develop spectrographs with the goal of achieving a precision better than about 15 m/s, a precision requested for the detection of gaseous giant planets. Among these different approaches, only a few were used in a systematic search to detect gaseous giant planets: in 1979, [12] introduced a HF absorption cell in front of the spectrograph to achieve internal precise wavelength calibration, while [13] designed an iodine cell for the same purpose.

The spectrum of stars of mass comparable or less than that of the sun have several thousand lines of absorption. [14] proposed to determine the radial velocity (velocity along the line of sight) by making a correlation of stellar spectrum and a template of the spectrum typical of this type of star. The position of the correlation peak gives access to the velocity of the star by concentrating the Doppler information of thousands of stellar lines. Figure 5 illustrates the wealth of information contained in 10% of the spectral domain used by the ESPRESSO spectrograph recently commissioned as part of the ESO Mount Paranal Observatory [15].

The development of photoelectric detectors allowed the realization of the first cross-correlation spectrographs [16, 17]. These precursor instruments demonstrated the enormous gain in efficiency (over 4000) of this technique compared to the old method based on photographic plates.

The development of CCD and optical fibers allowed the realization of cross-correlation spectrographs to determine the stellar velocity fluctuations gradually

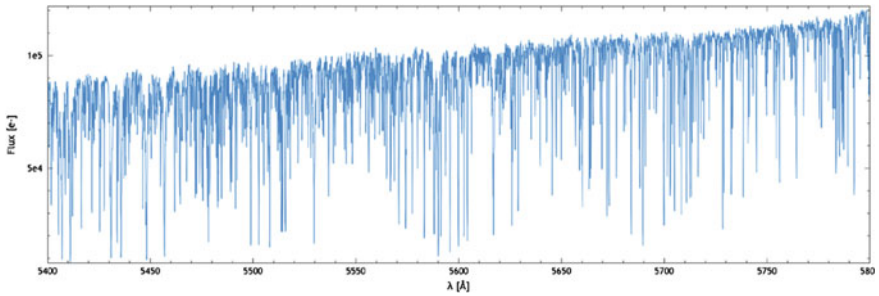

Fig. 5 Spectrum of HD 85,512 obtained with the ESPRESSO spectrograph installed at the ESO Paranal Observatory (Chile). Only 400 Å appear in the figure while the true spectral window of ESPRESSO is about 10 times larger. With ESPRESSO we can measure the important Doppler information contained in the spectra of solar-type stars or colder. The cross-correlation technique allows one to concentrate Doppler information from several thousand absorption lines to measure precise stellar radial velocities (Credit [15])

from 10–15 m/s in the 90's (ELODIE spectrograph, [18] to 1 m/s in 2003 (HARPS spectrograph, [19] then 0.1–0.2 m/s for ESPRESSO [15].

3 The Discovery of 51 Pegasi B

The context of the search for exoplanets at the beginning of the nineties is amazing. On the one hand, the groups involved in the search for exoplanets reach the conclusion that there are no gas giant planets (mass comparable to Jupiter) around solar type stars. These non-detections result from programs carried out over several years with comparable accuracies for all groups (about 15 m/s) but relatively small stellar samples [20, 21].

Moreover, the scenario to explain the formation of gas giant planets predicts that the orbital periods of such planets cannot be less than about ten years: A consequence of the necessity to agglomerate ice grains in the external regions of the accretion disks for the formation of planet of mass comparable to Jupiter [22].

Since spring 1994 Mayor and Queloz have initiated a program to search for low mass companions associated with solar-type stars. Our interest was as much on planetary companions (thus hoped with orbital periods of more than 10 years) or companions of brown dwarf's type. Brown dwarfs are objects of such low mass that nuclear reactions could not be initiated in their core (mass less than about 0.075 solar mass). Among the 142 stars of our sample, at the end of the first measurement season (beginning of 1995) the 12 measurements of the velocity of one star (51 Pegasi) presents a periodic variation of 4.2 days. Such a velocity could result from the influence of a possible planet with a mass greater than half that of Jupiter. Thus, an orbital period 1000 times shorter than expected from the formation scenarios! What could be the alternative causes of this variation?

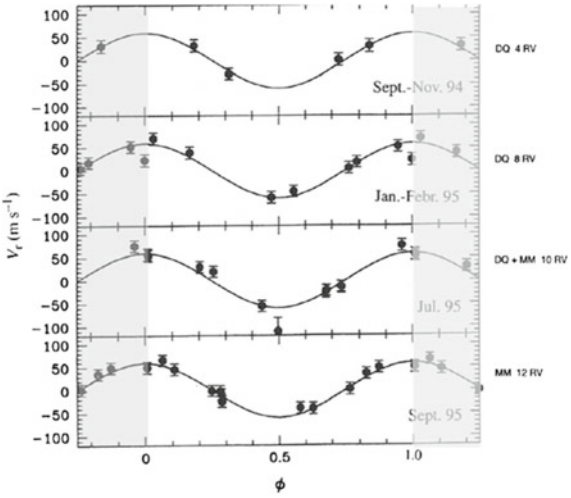

Fig. 6 Radial velocity measurements of 51 Pegasi obtained during four observing runs from September 1994 to September 1995. Adapted from [23]

The radial velocity only gives access to the minimum mass of the companion, corresponding to a line of sight in the plane of the orbit. Many stars exhibit pulsations. The atmospheres of stars of solar mass (or less) are affected by a magnetic activity at different levels of intensity. Such activity induces variations in the measured stellar velocity (jitter). The stability of the velocity variation observed during the following season (Fig. 6) as well as the very weak magnetic activity of 51 Peg, the impossibility of pulsation modes (gravity modes with periods of 4 days) and the photometric stability of the star convinced us that the influence of a Jovian planet was the most robust interpretation [23].

4 The Orbital Migration and the Observed Diversity of Planetary Systems

How to explain the presence of a giant planet so close to its star? 51 Pegasi b is located at 0.05 AU from its host star. Following the announcement of the discovery of this first Jovian planet (a hot Jupiter!) with a short orbital period. Lin et al. [24] show that such a period can be understood by considering the interaction of the planet and the accretion disk.

The planet interacts gravitationally with the accretion disk (before it dissolves). In return the excess density of the perturbation acts on the young planet and creates a very efficient orbital migration. This process was already studied for 15 years but had not been considered in the study of the mechanisms of planetary formation [25–28]. The numerous discoveries of extrasolar planets which followed this first detection demonstrated the importance of the phenomenon of orbital migration to understand the diversity of planetary systems. The discovery of this physical phenomenon has led

to a major change in our understanding of the mechanisms of system formation. If the basic idea of this interaction is simple, its understanding in the case of real systems is complex: too fast migration of protoplanets predicted by models, formation of multiple planetary systems leading to migrating protoplanets capturing each other in mean motion resonances, poorly known angular momentum transport and extraction mechanisms in the gaseous accretion disk, gravitational interactions between forming planets leading to scatterings and collisions, Kozai effect with distant stellar companions and many more (see Sect. 9).

Soon after the discovery of 51 Pegasi b, existing radial-velocity surveys of nearby solar-type stars were significantly expanded and new ones started with precision of 3–10 m/s [18, 29]. Additional hot-Jupiter discoveries quickly followed and over the next two decades, hundred giant exoplanets were found, spanning a wide range of masses and orbital distances.

In 2003, a new gain in precision was achieved with the construction of the "High Accuracy Radial Velocity Planet Searcher (HARPS) spectrograph at La Silla Observatory in Chile [19]. This fiber-fed vacuum spectrograph allows routine precision better than 1 m/s. A important number of systems with planets smaller than Neptune could be detected.

All the surveys carried out by different teams reveal the amazing diversity of planetary systems with orbital periods from few hours to dozens of years, eccentricities up to 0.93 and planetary masses from rocky planet range (about the Earth-mass) to several times the mass of Jupiter (more than one thousand Earth-mass), as illustrated by Fig. 7.

Planetary systems reveal very complex architectures frequently with resonant orbits. One of the most striking discovery made by these surveys was the observation

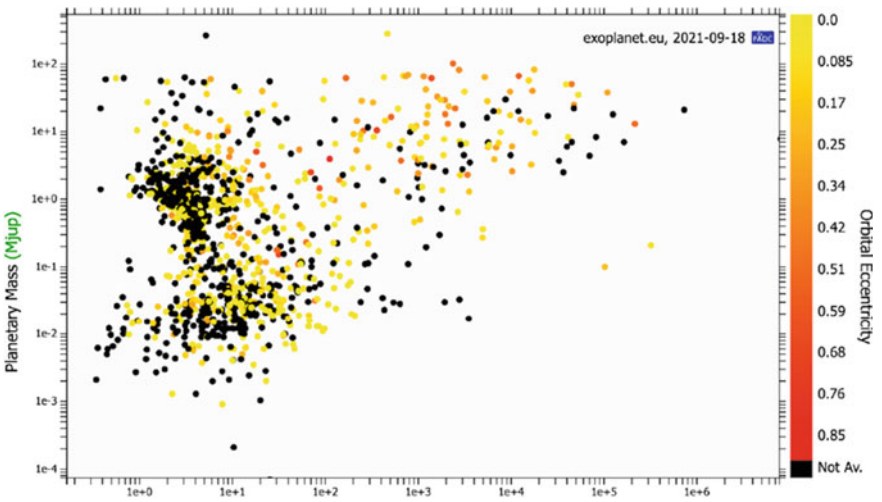

Fig. 7 Planetary mass versus the orbital period (in days) of observed exoplanets (from the exoplanet database "exoplanet.eu")

of an extremely rich population of planets with masses in the range of 1 to 10 Earth-mass (named "Super-Earths). That exoplanet population is dominant among the solar-type stars (with the present radial velocity precision), but that domain of mass is not present in the solar system.

5 Insights from Planetary Transit Searches

In addition to the enormous diversity of the architectures and orbital parameters of planetary systems, a remarkable diversity of internal structure of exoplanets has been revealed.

Soon after the discovery of the first transiting planet (Fig. 8; [30, 31]), several ground-based efforts and space missions have been developed to detect transiting planets.

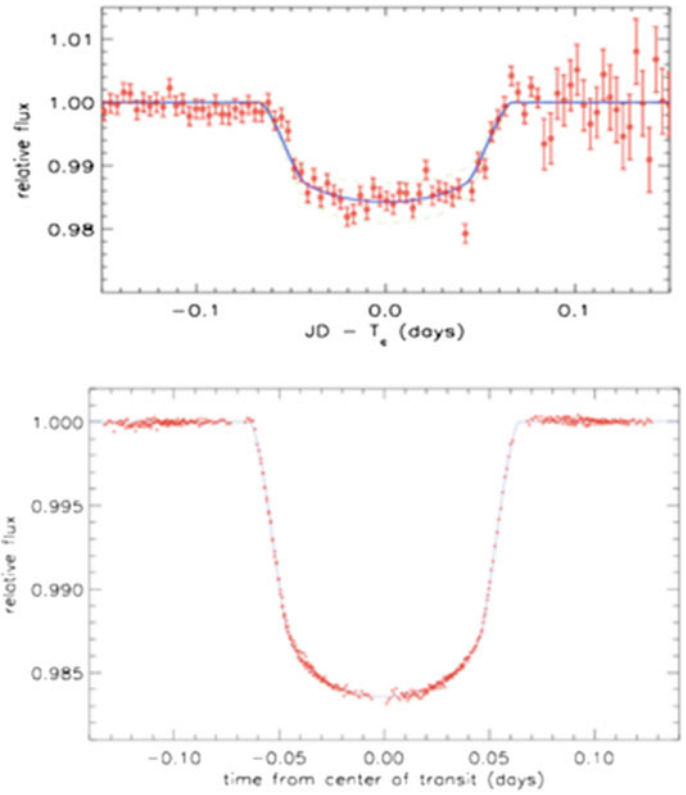

Fig. 8 September 9 and 16, 1999. A first planetary transit is observed, proving that Hot Jupiters are gaseous giant planets. The combination of observed mass and radius reveal a low mean density of the planet of 0.3 g/cm^3. **a** From [30]. **b** From [32]

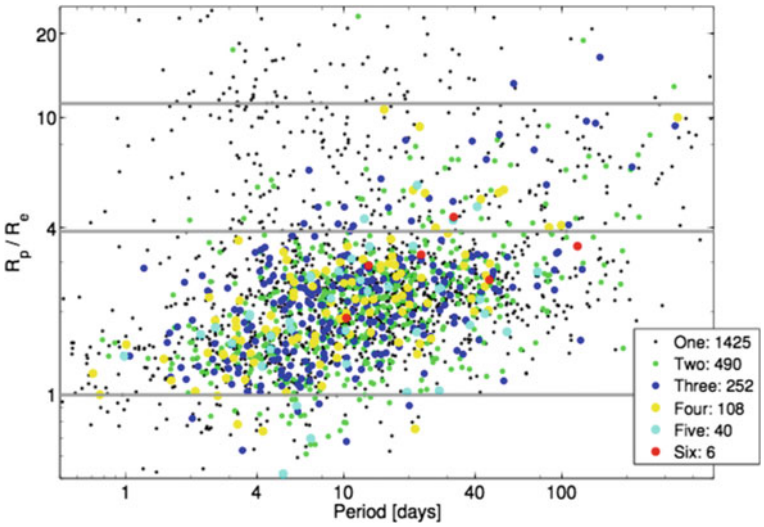

Fig. 9 Planetary transits observed by the Kepler mission The multiplanetary transits are coded with different colors -up to 6 planets (Credit [33])

The quality obtained outside the atmosphere shows the importance of space measurements. Many dedicated missions: CoRoT, Kepler, TESS, CHEOPS, and soon PLATO illustrate the great potential of this technique. The NASA Kepler mission has revealed thousands of planetary systems and multi-planetary systems (see Fig. 9).

The contrast of planetary transit permits the determination of the diameter of the planet relatively to the stellar diameter. But when the planet disappears behind the star, the decrease of the infrared luminosity gives access to the temperature of the planetary temperature. In some case the precise timing anomalies of multiplanetary transits provide an estimation of the planetary masses.

In addition, the transits offer an exceptional tool to get some information on the chemical composition of atmosphere of exoplanets: A possibility presented in following sections.

The space missions Kepler and Tess have detected several thousand planetary systems and the diameters of transiting planets discovered by these missions have been estimated. Few of them correspond to 1 to 3 Earth–radii and have been selected to determine their masses by Doppler spectroscopy and to explore the diversity of planetary structure into the R-M plane.

The planets with precise radii and masses smaller than 20 Earth-masses have been plotted on Fig. 10. The sequence of rocky planets is visible up to maybe 10 Earth-masses. A few planets with masses above 4–5 Earth-masses have an important atmosphere (mini-Neptunes having migrated or suffered from an important evaporation (see the section *Upper atmospheres*).

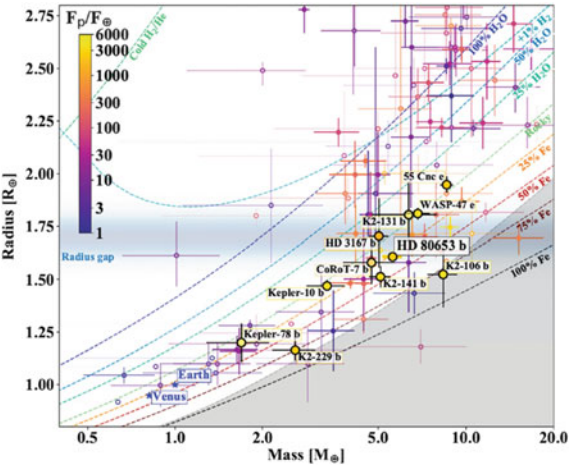

Fig. 10 Mass-radius diagram of planets smaller than 2.8 Earth radii [34]. The dashed lines show planetary interior models for different compositions as labeled [35]. Planets are color-coded according to the incident flux Fp, relative to the solar constant received on the Earth (for the full description see [34])

6 Planetary Atmospheres

Atmospheres of exoplanets are our main observing window into the physical and chemical phenomena occurring on these remote worlds. Atmospheric characterisation of exoplanets is booming as spectacular progresses have been achieved [36, 37], largely thanks to the increasing quality of available instrumentation [38]. Today, the most amenable targets for atmospheric characterisation are the transiting systems: during the transit, a minute fraction of the star light filters through the planet atmospheric limb. At wavelengths where atmospheric components absorb this light, the transit appears deeper than in white light. Observing a transit with a spectrograph thus allows one to measure the size of the planet as a function of wavelength, yielding the transmission spectrum of the planet atmosphere [39]. The atmospheric signal scales with the planet temperature, the mean molar mass of the atmosphere, the density of the planet and the size of the star. It is much easier to detect the primordial atmospheres of hot gas giants composed mainly of hydrogen and helium than that of rocky planets with secondary atmospheres. Among the latter category, rocky planets orbiting red dwarfs, the smallest and faintest main sequence stars, that are easier to characterise than their counterparts around Sun-like stars. This explains the huge interest for the planetary systems such as TRAPPIST-1 [40, 41], which features seven Earth-size planets, including three in the habitable zone [42, 43] of a star no larger than our Jupiter. This system is one of the key targets for the James Webb Space Telescope, which will attempt to determine the molecular composition of these potentially temperate atmospheres, twenty years after the Hubble Space Telescope provided the first spectroscopic signatures arising from the extreme atmosphere of a hot gas giant [44, 45]. Since then, the Hubble and Spitzer space telescopes has revealed the diversity of giant planet atmospheres [46] and how such planets, scorched by the intense stellar irradiation, are spectacularly losing their atmosphere

to space hinting at pathways for planetary evolution [47, 48] unthought of by just looking at our own Solar System.

Recently, ground-based observations with high-resolution spectrographs have enabled access not only to the chemical composition of exoplanet atmospheres but also to their dynamics, from measuring violent winds blowing at several kilometres per second [49] to mapping the condensation of iron across their surface [50]. In other words, we are on the verge of observing the (extreme) climates of exoplanets. Reaching out to more temperate planets and climates will benefit from the anticipated combination of high-resolution spectroscopy with high-contrast and high-angular resolution imaging. This is one of the biggest promises of the next generation of extremely large (up to 39 m!), which will receive their first light during the upcoming decade.

7 Upper Atmospheres

Many absorption features from a transiting exoplanet arise from atmospheric layers at low altitudes, where the increase in transit depth is proportional to the scale height (a measure of the atmospheric compactness). Optically thick clouds and hazes can hide the layers underneath, preventing the study of atmospheric properties even in hot gas giants with large scale height [46]. In Neptune-size planets and super-Earths, clouds generally preclude the detection of molecular features from the lower layers. Fortuitously these limitations disappear in the extended outer layers of hot exoplanets, called thermosphere and exosphere, which can yield absorption signals up to several percents.

This realization came with the seminal observation of the hot Jupiter HD209458b with the STIS spectrograph onboard the Hubble Space Telescope [45]. Transit spectroscopy in the stellar Lyman-α line, emitted in the ultraviolet (121.6 nm) by hot neutral hydrogen in the transition region and corona of a star, showed absorption by neutral hydrogen escaping from the upper planetary atmosphere. Indeed, the absorption of 15% of the stellar light by HD209458b, much deeper than absorption signatures from its lower atmosphere, can only be attributed to a layer of neutral hydrogen extending beyond the Roche lobe of the planet (the limit of its gravitational influence) into its exosphere. Analysis of this absorption signal showed that the planet is losing thousands of tons of gas per second into space and revealed a new regime of atmospheric escape unknown in the Solar System.

Theory backed by radiation-hydrodynamic atmospheric models suggests that strong deposition of high-energy X and ultraviolet (XUV) stellar radiation into close-in planets leads to a dramatic expansion of their upper atmosphere and its subsequent escape [51]. Theoretical studies of the exoplanet population and specific systems proposed that this “evaporation” could partly explain the lack of Neptune-size planets close to their star (see review by [52]). While hot Jupiters are too massive to erode substantially, hot Neptunes may be unable to retain their escaping atmosphere, evolving into rocky remnant cores [54, 55]. Understanding the origins and

evolution of exoplanets that orbit close to their star requires observing and modelling their upper atmospheres, to better constrain the evaporation process.

Observations of exospheres lie at a physically rich cross-road, connected to aeronomy, stellar physics and planetary evolution. Observing a transit with a spectrograph allows resolving absorption signals from exospheres as a function of wavelength. The spectral absorption profile then yields information about the density and velocity structure of the exosphere, as well as its interactions with the star. For example, interpretation of GJ436b Lyman-α spectra with a three-dimensional particle-based model showed that its exosphere is shaped by photoionization, stellar radiation pressure, and charge-exchange with the stellar wind [55]. Observations of exospheres can also be used to constrain the rate at which atmospheres lose mass, a critical input to understand their evolution. Interpretation of Lyman-α signals from hot Jupiters yielded mass loss rates of 10^8–10^{10} g/s (less than 0.1% of their mass per Gyr), confirming the stability of their atmosphere predicted by theory (e.g. [56]). In contrast the warm Neptunes GJ436b and GJ3470b, located at the border of the Neptunian desert (Fig. 11), could have lost up to 10–35% of their mass [57]. Did they form as larger planets that eroded down to Neptune-mass through evaporation? Or were they brought recently close to their stars without having time to lose much atmosphere? These questions show that the atmospheric evolution of exoplanets cannot be studied independently from their dynamical history. Until recently, measurements of atmospheric mass loss rates were limited by the fact that neutral hydrogen only probe exospheres—thus bringing a partial view on atmospheric escape—and that most transiting systems are too faint or distant for Lyα observations (the interstellar medium contains neutral hydrogen that completely occult stars beyond ~ 60 pc from Earth). The detection of many close-in planets transiting bright stars by recent ground

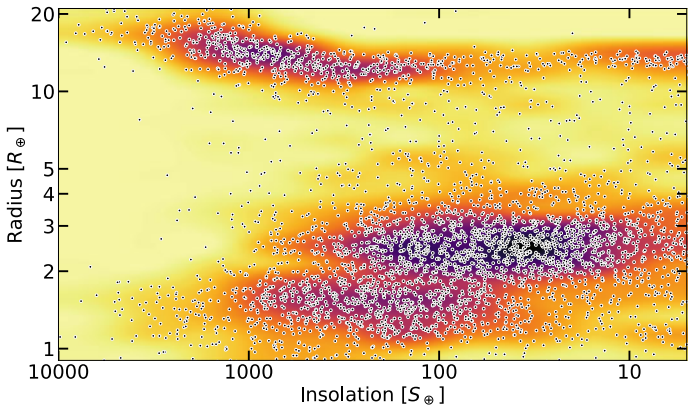

Fig. 11 Distribution of known planets (from exoplanets.eu), plotted as a function of their radius and insolation from their host star. Background is colored with planet occurrence, from low (yellow) to high (black). This representation highlights the non-uniform distribution of close-in planets, with features such as the desert of hot Neptunes

and space missions, such as TESS, together with the discovery of a new tracer of upper atmospheres, are now changing the picture.

For many years, the thermosphere of exoplanets remained unexplored observationally despite its importance to understand atmospheric escape. Indeed, this layer immediately below the exosphere is thought to be the birthplace of evaporation. The thermosphere of close-in planets absorbs such formidable amounts of XUV energy that the heated gas—mainly hydrogen—is predicted to expand hydrodynamically, in a global upward outflow, up to several planet radii, thus leading to much stronger mass loss rates than in the solar system. This scenario of hydrodynamical escape, which further predicts that heavy atoms are dragged upward by collisions with the expanding hydrogen outflow, is supported by the detection of oxygen, carbon, or silicon in exospheres (e.g. [58–60]). However, atmospheric expansion depends on a complex interplay between physico-chemical processes that are not well understood, such as the conversion of stellar energy into heating (e.g. [61]). Over the last decade many theoretical and numerical models were developed to simulate thermospheric expansion, with increasing complexity (e.g. [62]. Yet predictions for thermospheric structures still vary substantially because these models could not be validated against observations. High-resolution spectroscopy of hot giants in the visible succeeded in resolving the core of sodium lines up to the transition stratosphere/thermosphere [63]. The shape of these lines revealed the onset of the temperature rise predicted by the hydrodynamical expansion of the thermosphere, further showing the need for direct observations of this region to validate theories. This eventually became possible thanks to the lines of neutral helium at 1.08 μm, which arise from a metastable level that can be heavily populated within the thermosphere [64]. These lines are not absorbed by the interstellar medium like the Ly-α line and are not as sensitive to stellar activity as other tracers. Helium was first detected in a hot Jupiter [65] through low-resolution transit spectroscopy with the WFC3 instrument on the Hubble Space Telescope, which however prevented resolving the absorption lines and determining their layer of origin (Fig. 12). The servicing of the near-infrared, high-resolution spectrograph CARMENES then allowed resolving the metastable helium lines in a warm Neptune [66], revealing that they arise from a thermosphere expanding up to the Roche Lobe and swept by zonal winds. Detections of extended thermospheres were further obtained with CARMENES for other hot Jupiters (eg [67–69], confirming the helium triplet as the missing observational link between close-in planets stratosphere and exosphere. The spectral identification of helium lines with high-resolution ground-based spectrographs (CARMENES, SPIrou, NIRPS) opens a key window into atmospheric escape for a wide variety of planets.

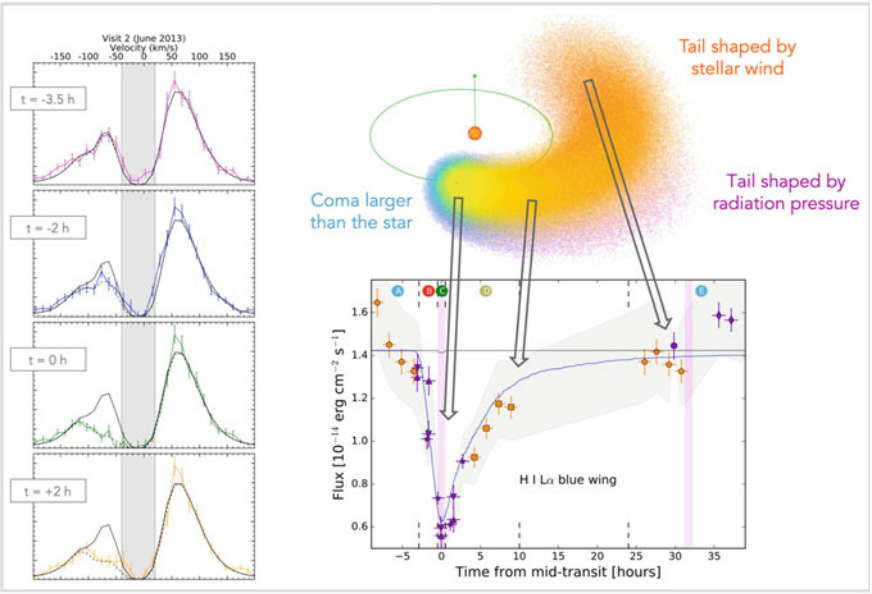

Fig. 12 Exosphere of the warm Neptune GJ436b. Left panels show the stellar Lyman-α line observed with *HST/STIS* at different times of the transit (the time relative to the center of the transit is shown in each panel), clearly highlighting the absorption by neutral hydrogen in the planet exosphere. The Lyman-α core is lost to interstellar medium absorption (grey band). The bottom right panel show the integrated absorption signal as a function of time, with a model light curve from particle-based simulations in red. The optical transit of the planet is shown as a black line. The corresponding simulation of the exospheric tail and coma is shown in the top right panel. (adapted from [55])

8 Planet Formation

Observations of the extrasolar planets and of the solar system planets provide the observational constraints upon which today's theory of planet formation and evolution is built. For extrasolar planets, the knowledge about an individual planet or planetary system is often limited (even thought that this is currently changing with an increasing number of exoplanets where measurements of the bulk and atmospheric composition become available). But given the number of now a few thousand exoplanets, the exoplanets can be treated as a population which put statistical constraints on formation models [70–72]. The solar system, on the other hand, yields very detailed observational constraints for formation theory [73, 74], but for only one specific planetary system which might not be representative.

These observational constraints are necessary, because constructing a successful theory of planet formation based on the first principles of physics only has proven to be difficult. The detection of 51 Pegasi b as a giant planet orbiting extremely close to its host star was, for example, not anticipated by most theoretical models [75,

76], even though that it was in principle known in the planet formation literature well before the detection that there are physical processes which could move giant planets very close to the star (namely the exchange of angular moment between a protoplanet and its nascent gaseous disk leading to orbital migration, [25, 26, 28]. The difficulties that the theory of planet formation faces are multifold: firstly, it must explain the growth from micrometer-sized dust grains to giant planets with a radius of about 100′000 km. At each size scale, different physical processes are of import. Second, it must follow the formation and evolution of a planetary system over millions or even billions of dynamical timescales, making a direct simulation in the computer difficult. Third, the formation involves strong non-linearities, back-reactions (for example from the gaseous disk on an embedded planet and vice versa), and decisive instabilities. Finally, a multitude of physical theories play a role, like gravity, hydrodynamics, thermodynamics, radiative transfer, magnetic fields, and high-pressure physics. Additionally, for practical reasons it is not possible to study most of the astrophysical processes in a laboratory, except for the very early growth phases involving only tiny dust grains.

Nevertheless, the wealth of observational constraints from the exoplanets and the solar system together with an ever-growing number of theoretical studies has allowed to construct the modern paradigm of planet formation [77]. To keep the problem practicable, the simplifying assumption is made that the process of planet growth can be spilt in several successive growth phases from small to large. It is clear that in reality, several phases will partially run in parallel, causing more complex formation pathways. A schematic, much simplified representation of the current picture of planet formation is shown in Fig. 13.

Everything starts with the formation of star, setting the starting time $t = 0$. Collapsing gaseous clouds in star forming regions (top left corner) rotate slowly. During their collapse under their own gravity, the rotation speeds up because of angular momentum conservation. Thus, not all infalling matter can directly fall on the forming star, but rather forms a surrounding protostellar disk [78]. Initially, the mass of the disk is comparable to the mass of the star, and violent gravitational instabilities may occur [79]. After a timescale of about 10^5 years, most of the gas infall from the cloud on the star and disk ends. The disk then enters a longer evolutionary phase (so called Class I and II) of about 10^6 to 10^7 years [80, 81]. These disks are now called protoplanetary disks. Four interferometric images obtained with ALMA sub-mm observations of actual disks [82] are shown in the bottom left corner of Fig. 13. These Class I and II disks, consisting mainly of hydrogen and helium gas and about 1% dust, are the formation environment of the planets. In the classical picture of viscous accretion disks [83, 84], these disks are slowly accreted onto the star because of turbulent angular momentum redistribution.

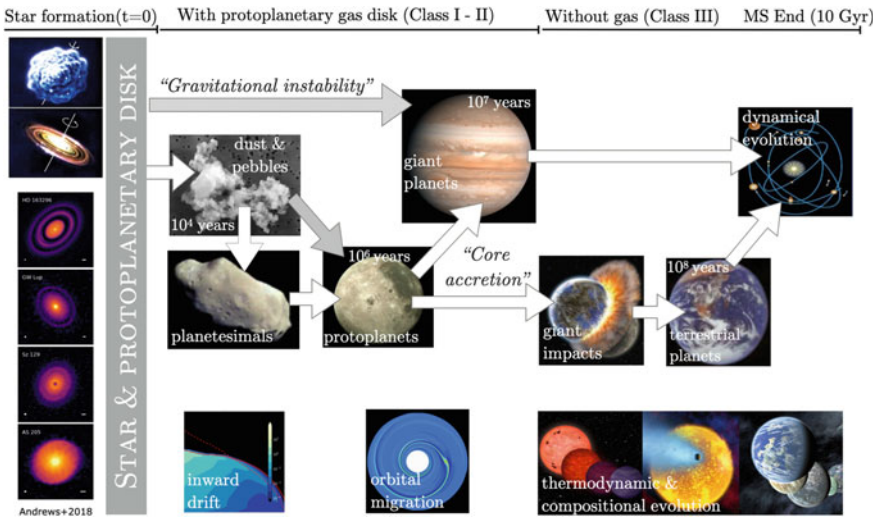

Fig. 13 Sketch of the sequential paradigm of planet formation. At the top, the main temporal phases (from the moment of star formation to the end of the main sequence (MS) of a star) are shown. In the middle, the different stages of growth and evolution are represented. White arrows show the classical picture while gray arrows show alternative hypotheses. At the bottom, physical processes of key importance at the different growth phase are show (Credit: W. Benz/C. Mordasini. Sub-figures from [82, 85, 86], NASA/ESA, E.K. Jessberger, D. Brownlee, Alfred Vidal-Madjar, Lynette R. Cook)

The most widely accepted formation theory is the so-called core accretion paradigm,[1] the stages of which are shown in the middle of the figure. It is a multi-staged bottom-up process. The very first growth stage regards the initially tiny dust grains of micron size that were accreted into the forming disk together with the infalling gas. On a short timescale (about 10^4 years in the inner disk), these dust grains grow by pairwise coagulation and form larger, about cm-sized objects called pebbles [89]. These pebbles are subject to a process known as inward drift [90]: unlike the disk gas, the solids do not feel a pressure support, implying that they need to orbit the star a bit faster than the gas to compensate the star's gravitational attraction. They thus feel a headwind, which in turn causes them to spiral inward towards the host star. This inward drift becomes extremely fast (a few years duration only) for about meter-sized bodies, threatening the very survival of the pebbles. For years, this difficulty has been known as the so-called meter barrier. The inward drift is not even the only problem at the 1 m scale: at this scale, the mutual collision also become so energetic that the bodies rather shatter upon contact rather that they grow [91].

[1] The term core accretion is sometimes used in a more restricted sense for the formation of giant planets only [87, 88]. Here we use it to describe the entire bottom-up growth process starting from tiny solids (dust).

The modern solution to this problem is that the object size directly jumps from pebbles to planetesimals with sizes of 1 to about 100 km, avoiding the difficult meter scale. For bodies of km-size, the inward drift is no longer of big concern. This jump is possible thanks to the so-called streaming instability [92]. It is a gas–solid instability that concentrates the pebbles locally because of the drag they experience from the disk gas. This concentration can reach a degree that the concentrated pebbles cloud eventually collapses under their own gravity forming directly large planetesimals [93].

In the next step, which is the growth from planetesimals to protoplanets of a size of about 1000 km, gravity becomes the governing force relative to drag. Growth at this stage proceeds via two-body collisions. Because of a mechanism known as runaway growth [94], larger protoplanets grow faster than smaller ones. This splits the population of bodies into smaller background planetesimals and larger protoplanets (also known as planetary embryos). These protoplanets also collide mutually in giant impacts. Once these protoplanets reach a mass on the order of about 0.1–1 Earth mass, their gravitational interaction with the gas disk becomes important and typically results in the inward migration of the bodies. Like the drift of pebbles at small scales, inward migration can become so fast that there is a danger that all protoplanets migrate into their host star [95, 96]. The mechanisms slowing down inward migration are not yet fully understood but may involve inhomogeneities in the disk [97], non-isothermal effects [98, 99] or a disk structure caused by magnetohydrodynamic winds [100].

Most protoplanets remain of a rather limited mass (a fraction of an Earth mass to a few Earth masses) during the presence of the gas disk because of the limited availability of building blocks in the inner disk [101] or because of the long timescale of collisional growth in the outer disk [102]. Some protoplanets however, especially those forming just outside the iceline, may manage to grow to masses of about 10 Earth masses during the lifetime of the gaseous disk. The iceline is the orbital distance from the star outside of which the temperature in the disk is sufficiently cold for water to condense. Thus, there are more building blocks available compared to the inner disk where only refractory materials (mainly silicate and iron) can condense. Classical disk models [103, 104]) found an iceline position of about 2–3 AU, i.e., between the terrestrial planets and the gas giants.

At a (core) mass of about 10 Earth masses, another effect becomes important, namely that the protoplanet now becomes able to efficiently accrete gas from the disk, surrounding itself with a gaseous envelope [105, 106]. While also protoplanets of lower mass are able to bind some nebular gas, the accretion timescale of the gas, which is regulated by the ability of the planet's gaseous envelope to cool and contract [107], becomes shorter than or comparable to the disk lifetime when this "critical core mass" [87, 88] is reached. In this case, planetary gas accretion can become very fast ("gas runaway", [108] and the planet quickly increases in mass to become a giant planet. Gas accretion necessarily stops when the nascent gas disk vanishes, which happens after a few to 10 million years [109].

Most protoplanets, however, do not become massive enough to accrete a lot of gas. After the dissipation of the gas, in the Class III disk stage when the disks consist only of leftover small solids and the newly formed planets, these protoplanets

start to mutually excite their eccentricities and inclination because of gravitational interactions. During the presence of the gas disk, the eccentricities and inclination are in contrast damped by the nebular gas. This excitation eventually leads to giant impacts and further growth. In the inner planetary system, this eventually leads to the emergence of terrestrial planets on a timescale of about 100 million years. For Earth, a biproduct of this stage is the formation of the moon in an off-axis giant impact [110]. Finally, over long timescales comparable to the system age (Gigayears for most stars), all bodies in the system further interact gravitationally and settle into a dynamical architecture that is quasi-stable over these long timescales [111]. It is in this context interesting to note that while apparently stable over long times, it is still possible for some solar system planets (Mercury, Venus) to become unstable and collide at some point in the far future [112].

During this final (and longest) phase, not only the planetary orbits and masses still evolve further. The planets also evolve individually thermodynamically and compositionally: giant planets cool and contract [113]; while close-in intermediate mass planets (Neptunian and sub-Neptunian) undergo important mass loss because of the atmospheric escape of their H/He envelope [47, 48]. Terrestrial planets finally undergo geophysical cycles, plate tectonics, and various interior-atmosphere interactions like outgassing [114]. This sets the astro- and geophysical context in which life may emerge and evolve.

In Fig. 14, gray arrows indicate potential alternative pathways relative to this classical picture. The existence of such alternatives shows that even on a sketch level, our understanding of planet formation is far from complete or definitive. The first such arrow directly connects dust/pebbles and protoplanets. It reflects that protoplanets may also grow from directly accreting cm-sized pebbles instead of km-sized planetesimals (so-called pebble accretion, [115]. Especially beyond the iceline, pebble accretion makes fast growth possible because gas drag enhances the capture radius of the protoplanets. This speed us the formation of giant planet cores [116].

However, even when considering the potential speed-up offered by pebbles, growing giant planet cores (i.e., reaching a critical core mass of 10 Earth masses before the gas disk dissipates) remains very challenging at orbital distances beyond about 10 AU [117]. Given that giant planets have been detected at orbital distances far beyond this limit (e.g., around HR 8799, [118], this motivates a completely different formation pathway for giant planets, namely gravitational instability [119– 121]. In contrast to core accretion, which is a bottom-up process, this is a top-down mechanism. In the gravitational instability mechanism (also known as direct collapse model) it is thought that large patches of the protoplanetary disk collapse under their own gravity and directly form large clumps of gas, a process that is most efficient in the outer disk [122]. These clumps can then further collapse to form giant planets [123]. Disks need to be massive for gravitational instabilities to occur, which is the case during the early infall phase [124]. Whether such disk fragmentation in the end really leads to the formation of giant planets also, and not merely more massive objects like brown dwarfs or binary stars, is still a matter of debate [125].

Given this sequential picture of planet formation, attempts were made to construct so called global end-to-end models of planet formation and evolution [126, 127].

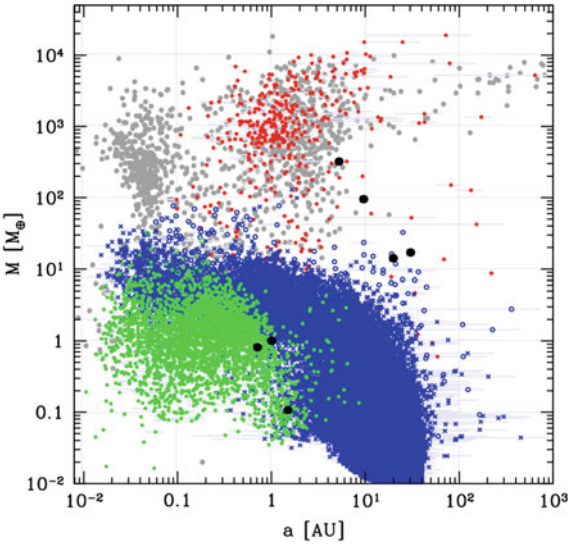

Fig. 14 Synthetic mass-distance diagram predicted by a global planet formation and evolution model. The coloured points shows the planets predicted by the model in synthetic systems of 1000 solar-like stars at an age of 5 Gyr. Green, blue, and red points represent silicate-iron, water-rich, and gas-dominated planets, respectively. Black and gray dots show the solar system and the exoplanets. The model assumes initially hundred lunar-mass proto-planets to be present in each disc (adapted from Emsenhuber et al. [128])

These models combine simplified descriptions of the aforementioned stages and physical processes into one large global model [129]. This allows these models to directly predict (given initial conditions which are the properties of the parent protoplanetary disk) the finally resulting planetary system. Initial such attempts employed strongly simplified formation models, relying for example on the assumption that only one planet forms per disk. Modern models have in contrast reached a considerable degree of complexity [130] and take for example the gravitational interaction of many concurrently forming protoplanets into account [131]. Such global models make statistical comparisons between theoretical predictions and exoplanet demographical observations possible by means of planetary population synthesis [126, 132, 133]. In this method, the initial conditions of a global model are varied in a Monte Carlo way according to distributions of the disk properties like initial disk gas and solid mass. These distributions are derived from observations of protoplanetary disks [82, 91, 134].

Figure 14 shows an example of the results of a population synthesis simulation of the formation of multiplanetary systems, considering the principal various phenomena evoked [128]. This allows a statistical comparison of the observations in the various quantities (in the figure planetary mass, orbital distance, orbital eccentricity, and bulk content of chemical elements—for reviews, cf. [133, 135]). Modern population syntheses can reproduce [128] several observed key statistical features

of the exoplanet population [70]. As the global models rely on the results of many specialized models (for the accretion of the solids and the gas, for orbital migration, for gravitational interactions etc.), these results give a certain impression of the current state and maturity of the field of planet formation theory. These results indicate that important progress has been made in our understanding of planet formation thanks to the statistical constraints offered by the exoplanet population. However, the global models fully rely on the specialized sub-models that they combine, and on this level, there are still many poorly understood aspects. Important examples, which represent areas of current active research, include the relative importance of growth via planetesimals versus growth via pebbles [136], or the angular momentum transport/extraction mechanism responsible for the temporal evolution of the protoplanetary disk (turbulent viscosity versus magnetohydrodynamic disk winds [137]. Future observations of both protoplanetary disks and extrasolar planets, including actively accreting planets observe during the formation process (like PDS70b, [138]), will help to improve our understanding of these processes.

9 The Search for Life

In the beginning of this chapter, we mentioned the idea of the plurality of the worlds and this quote from Epicurus (341–270 BC) which read, "*Worlds are in an infinite number some of them similar to our own one, some others being different…living species, plants and all the other visible things could exist in some worlds and could not in others.*". Not only Epicurus was speculating on the existence of other worlds but also on their habitability: some planets (worlds) should be habitable and inhabited, some of them not. With the discovery of planets orbiting other stars than our own, we now know that other worlds are indeed very common. However, we have not yet been able to identify another inhabited world inside or outside our solar system. This brings us to this concept of "habitability". What do we mean by "habitable"?

Following Epicurus, a "habitable" world would be a planet which hosts life, which brings us to the first definition of a habitable planet:

A habitable planet is a planet that meets all the necessary conditions for the emergence and development of life.

This definition is a bit tricky as there is also no consensus on the actual definition of life. Would we be able to recognize a form of life different than what we know on Earth? Probably not. Which leads us to the second definition of life:

A habitable planet is a planet that meets all the necessary conditions for the emergence and development of life as we know it on Earth.

This definition is a bit more precise and is going to help us go farther in our description of habitability, because we know life on Earth, we have been studying it for centuries. Based on this knowledge, we know that life as we know it on Earth needs:

(1) **Building blocks**. Those are chemical elements which are in the molecules which constitute any life form: carbon, hydrogen, oxygen, phosphorous and sulfur (or CHNOPS);
(2) **Liquid water**. It is one of the best solvents for organic chemistry. It allows for hydrogen bonds which help stabilize big organic macromolecules. Liquid water is also stable for a wide range of pressure and temperature.
(3) **A source of energy**. This source of energy can be the Sun for the "photosynthetic life" at the surface of the Earth, but it can also be energy obtained by the oxidation of inorganic molecules like iron or magnesium for the "chemotrophic life" on the Earth's ocean floor.
(4) **A surface**. We do not expect life as we know it on Earth to be able to exist on a planet like Jupiter. A surface is also a way to ensure the building blocks are available in numbers.

This leads us to the third definition of habitability:

A habitable planet is a planet that has a surface, liquid water, carbon (e.g. in the form of organic molecules, with other nutrients such as H, N, O, P, S) and a source of energy (e.g. photons).

One of the most challenging condition to fill of this list is the presence of liquid water. Furthermore, to be able to detect life, it is better to have surface liquid water. First, Earth is the only body of the solar system which can sustain surface liquid water and incidentally is the only place we know to host life. Second, bodies with subsurface water (such as Europa) do not constitute the best targets to look for forms of life: they are difficult to explore in situ (drilling might be necessary) and extremely difficult to explore remotely. Third, at the surface, life can directly use the energy brought by solar photons. And fourth, a life evolving at the surface can more easily modify the atmosphere in a detectable way, which is an extremely useful factor when the observer is located very far away.

This leads us to the final definition, which is the one adopted by the community of exoplanet scientists:

A habitable planet (or moon) is a planet (or moon) that has surface liquid water.

This definition is the one that motivated the concept of the Habitable Zone [139]. The exoplanet community is therefore on the hunt of small rocky planets in the habitable zone of their star: these planets have a surface, potentially with liquid water on top of it, and potentially hosts a form of life which could have significantly impacted the atmosphere.

About 200 planets have been detected in the HZ (http://hzgallery.org/), but most of them are gaseous planets. However, we now count a few rocky planets in this list. As of January 2022, we have two promising candidates to search for life: Proxima-b [140] and the TRAPPIST-1 planets [41]. Proxima-b has a mass of at least 1.2 Earth mass and is thought to be rocky. TRAPPIST-1 hosts 7 planets very similar to Earth in terms of mass and radius and 3 of them are in the HZ. Both stars, Proxima-Centauri and TRAPPIST-1 are very low mass stars (12% and 9% the mass of the Sun respectively).

However, it is a big step to go from identifying a rocky planet in the HZ to identifying an inhabited planet. And this leads us to the concept of biosignatures. Biosignatures are signatures of life, so any signal or combination of signals which could be interpreted as having a biological origin. For instance, for the Earth, dioxygen is a biosignature as it is coming from organisms whose metabolism relies on photosynthesis. Dioxygen is thus sometimes proposed as a biosignature. Other possible biosignatures are methane (CH_4), nitrous oxide (N_2O) and sulfur dioxide (SO_2). The challenge that biosignatures present is the risk of false positives. For instance, due to the high activity of low-mass stars, the planet loses its water over the course of its evolution (via evaporation, see previous section). This abiotic process leads to a dioxygen build-up in the atmosphere. This process is thought to be important for planets like Proxima-b and the TRAPPIST-1 planets. This means that even if we detect only O_2 in their atmosphere, we will not be able to say that it is because of the existence of life. All the other possible biosignatures suffer from this issue, with geochemical or photochemical processes, which can produce these gases (see [141] for a discussion about the potential of O_2 as a biosignature, accounting for its false positives and proposing observing strategies).

It is an ongoing effort of the community to unravel the astronomical, geophysical, chemical context of the planets to understand these false-positive producing processes. This will allow the community to check if a signature could be from a form of life on the planet: if no other abiotic process (geophysical, photochemical, …) can explain this signature, we might have found life. Other possible ways to detect life would be to observe clearly artificial signals (as what SETI tries to accomplish) or to detect fractional circular polarization in reflected light (see [142]. Such polarization is a unique signature of the homochirality of the molecules, macromolecules and biomolecular architectures which make up life.

Another difficulty is the observation of the weak signals from which these signatures can be obtained. It requires to detect chemical species in the atmosphere of exoplanets. As explained earlier in this chapter, transit spectroscopy will allow us to probe the atmosphere of small rocky planets in the HZ such as the TRAPPIST-1 planets. For instance, with the James Webb Space Telescope, it could be feasible to detect CO_2 for the fifth planet of TRAPPIST-1 [143], which is also the planet with the highest probability of hosting surface liquid water [144]. Unfortunately, Proxima-b does not transit, but its atmosphere can be probed using high-contrast imaging and high-resolution spectroscopy [145]. The idea is to isolate the light being reflected by the planet and feed this light in a spectrograph. Analyzing the reflected light can inform on the composition of the atmosphere. [146] have shown that it could be possible to detect dioxygen in the atmosphere of Proxima-b with this technique. This led to the RISTRETTO instrument being developed. RISTRETTO will be installed at the Very Large Telescope in Chile in the late 2020s and one of its goal is to characterize the atmosphere of Proxima-b. It is not the only one and there are other instruments are being developed to characterize the atmospheres of planets this way (like HIRES at the European-Extremely Large Telescope).

Even if the community appears still a long way from being able to detect biosignatures, with the precision required to be able to claim a detection of life, tremendous

efforts are being done in that direction with ingenious ideas to circumvent the biggest hurdles. The community is working on designing a new generation of instruments which might detect the presence of a set of molecules with information on their relative abundances. Also, we need instruments that will be able to not only characterize the HZ planets of low-mass stars like TRAPPIST-1 or Proxima-Centauri but Earth-like planets orbiting in the HZ of Sun-like stars. Indeed, planets around low-mass stars have a very different environment compared to our Earth. Low-mass stars are very active stars which can bombard planets with huge amounts of Ultraviolet radiations, which could damage surface life (e.g., [147]. Among future mission's concepts which aims at detecting biosignatures in the atmosphere of an Earth-like planet around a Sun-like star, there is the Large Interferometer For Exoplanets (LIFE) mission concept [148], and the Starshade rendezvous mission concept [149].

Acknowledgements This work has been carried out in the frame of the National Centre for Competence in Research PlanetS supported by the Swiss National Science Foundation (SNSF).

References

1. P. Connes, *History of the Plurality of Worlds. The Myths of Extraterrrestrials Through the Ages*, ed. by J. Lequeux (Springer Historical & Cultural Astronomy Series, 2020). ISBN: 978–3–030–41448–1. https://doi.org/10.1007/978-3-030-41448-1
2. S.J. Dick, From the physical world to the biological universe: historical developments underlying SETI bioastronomy. *The Search for Extraterrestrial Life-The Exploration Broadens Lecture Notes in Physics*, ed. by J.Heidmann, M.J. Klein, vol. 390, Conference Proceedings. Bioastronomy (Springer, Berlin, Heidelberg, 1991), pp. 356–363. ISBN: 978–3–540–54752–5
3. O. Struve, Proposal for a project of high-precision stellar radial velocity work. Observatory **72**, 199–200 (1952)
4. S.V.W. Beckwith, A.I. Sargent, R.S. Chini et al., A survey for circumstellar disks around young stellar objects. A J **99**, 924–945 (1990)
5. M.J. McCaughrean, C.R. O'Dell, Direct imaging of circumstellar disks in the Orion Nebula. A. J. vol. 111, n.5, 1977 (Published online by Cambridge University Press, 1996), 12 Apr. 2016. https://doi.org/10.1086/117934
6. K. Strand, 61 Cygni as a triple system. Publ. Astron. Soc. Pacif. **55**, 322, 29–32 (1943). https://www.jstor.org/stable/40670169
7. S.L. Lippincott, The unseen companion of the fourth nearest star, Lalande 21185. A.J. **65**, 349L (1960)
8. P. Van de Kamp, Astrometric study of Barnard's star from plates taken with the 24-inch Sproul refractor. A.J **68**, 515–521 (1963).https://doi.org/10.1086/109001
9. S.H. Pravdo, S.B. Shaklan, An ultracool star's candidate planet. ApJ **700**, 623–632 (2009). https://iopscience.iop.org/article/10.1088/0004-637X/700/1/623
10. M. Perryman, J. Hartman, G.Á. Bakos, L. Lindegren, Astrometric exoplanet detection with Gaia. ApJ **797**, 14. 22pp. (2014). https://doi.org/10.1088/0004-637X/797/1/14
11. D. Belorizky, Le soleil, étoile variable. L'Astronomie **52**, 359–361 (1938)
12. B. Campbell, G.A.H. Walker, Precision radial velocities with an absorption cell. PASP **91**, 540–545 (1979)
13. G.W. Marcy, R.P. Butler, Precision radial velocities with an iodine absorption cell. PASP **104**, 270–277 (1992). https://iopscience.iop.org/article/10.1086/132989

14. P. Fellgett, A proposal for a radial velocity photometer. Opt. Acta: Int. J. Opt. **2**(1), 9–16 (1955). https://doi.org/10.1080/713820996
15. F. Pepe, S. Cristiani, R. Rebolo et al., ESPRESSO at VLT. On-sky performance and first results. A&A **645**, A96. 26pp. (2021). https://doi.org/10.1051/0004-6361/202038306
16. A. Baranne, M. Mayor, J.L. Poncet, CORAVEL: a new tool for radial velocity measurements. Vistas Astron. **23**, 279 (1979). https://doi.org/10.1016/0083-6656(79)90016-3
17. R.F. Griffin, A photoelectric radial-velocity spectrometer. ApJ. **148**, 465–476 (1967). https://ui.adsabs.harvard.edu/abs/1967ApJ...148..465G/abstract
18. A. Baranne, D. Queloz, M. Mayor et al., ELODIE: a spectrograph for accurate radial velocity measurements. Astron. Astrophys. Suppl. Ser. **119**, 373–390 (1996). https://doi.org/10.1051/aas:1996251
19. M. Mayor, F. Pepe, D. Queloz et al., Setting new standards with HARPS. Messenger **114**, 20–24 (2003). ISSN0722–6691
20. G.W. Marcy, R.P. Butler, Brown dwarfs and planets around solar-type stars: searches by precise velocities, in *The Bottom of the Main Sequence—and Beyond. Proceedings of the ESO Workshop Held in Garching*, Germany, 10–12 Aug. 1994, ed. by Christopher G. Tinney. Part of the ESO Astrophysics Symposia, Book Series (Springer-Verlag Berlin Heidelberg New York, 1995), pp. 98–108 (1994)
21. G.A.H. Walker, A.R. Walker, A.W. Irwin et al., A search for Jupiter-mass companions to nearby stars. Icarus **116**, 359–375 (1995). https://doi.org/10.1006/icar.1995.1130
22. A.P. Boss, Proximity Jupiter-Like planets to low-mass stars. Science **267**(5196), 360–362 (1995). https://doi.org/10.1126/science.267.5196.360
23. M. Mayor, D. Queloz, A Jupiter-mass companion to a solar-type star". Nature (London) **378**, 355–359 (1995). https://doi.org/10.1038/378355a0
24. D.N.C. Lin, P. Bodenheimer, D.C. Richardson, Orbital migration of the planetary companion of 51 Pegasi to its present location. Nature **380**, 606–607 (1996). https://doi.org/10.1038/380606a0
25. P. Goldreich, S. Tremaine, Disk-satellite interactions. ApJ. **241**, 425–441 (1980). https://doi.org/10.1086/158356
26. D.N.C. Lin, J. Papaloizou, On the tidal interaction between protoplanets and the protoplanetary disk. III—Orbital migration of protoplanets. ApJ **309**, 846–857 (1986)
27. J. Papaloizou, D.N.C. Lin, On the tidal interaction between protoplanets and the primordial solar nebula. I—Linear calculation of the role of angular exchange. ApJ **285**, 818–834 (1984)
28. W.R. Ward, Density waves in the solar nebula: differential Lindblad torque. Icarus **67**, 164–180 (1986). https://doi.org/10.1016/0019-1035(86)90182-X
29. R.P. Butler, G.W. Marcy, A planet orbiting 47 Ursae Majoris. ApJL **464**, L153–L156 (1996)
30. D. Charbonneau, T.M. Brown, D.W. Latham, M. Mayor, Detection of planetary transits across a Sun-like star. ApJ **529**, L45–L48 (2000)
31. G.W. Henry, G.W. Marcy, R.P. Butler, S. Vogt, A transiting "51Peg-like" planet. ApJ **529**, L41–L44 (2000). https://doi.org/10.1086/312458
32. T.M. Brown, D. Charbonneau, R.L. Gilliland et al., Hubble Space Telescope time-series photometry of the transiting planet of HD 209458. ApJ **552**, 699–709 (2001)
33. N.M. Batalha, J.F. Rowe, S.T. Bryson et al., Planetary Candidates Observed by Kepler. III. Analysis of the first 16 months of data. ApJSuppl. **204**(2), 24 21 pp. (2013). https://doi.org/10.1088/0067-0049/204/2/24
34. G. Frustagli, E. Poretti, T. Milbourne et al., An ultra-short period rocky super-Earth orbiting the G2-star HD 80653. A&A, **633**, A133, 11 pp. (2020). https://doi.org/10.1051/0004-6361/201936689
35. L. Zeng, S.B. Jacobsen, D.D. Sasselov et al., Growth model interpretation of planet size distribution. Proc. Natl. Acad. Sci. USA **116**, 9723–9728 (2019). https://doi.org/10.1073/pnas.1812905116
36. A.S. Burrows, Highlights in the study of exoplanet atmospheres. Nature **513**, 345–352 (2014). https://doi.org/10.1038/nature13782

37. N. Madhusudhan, *ExoFrontiers: Big Questions in Exoplanetary Science*, ed. by N. Madhusudhan. 306 pp. Oct. 2021 (Copyright © IOP Publishing Ltd 2021, 2021). IOP ebook. Online ISBN: 978–0–7503–1472–5; Print ISBN: 978–0–7503–1470–1. https://doi.org/10.1088/2514-3433/abfa8f
38. F. Pepe, D. Ehrenreich, M. Meyer, Instrumentation for the detection and characterization of exoplanets. Nature **513**, 358–366 (2014). https://doi.org/10.1038/nature13784
39. S. Seager, D. Sasselov, Theoretical transmission spectra during extrtasolar planet transits. ApJ **537**(2), 916–921 (2000)
40. M. Gillon, E. Jehin, S.M. Lederer et al., Temperate Earth-sized planets transiting a nearby ultracool dwarf star. Nature **533**, 221–224 (2016). https://doi.org/10.1038/nature17448
41. M. Gillon, A.H.M.J. Triaud, B.-O. Demory et al., Demory, B.O. 2017, Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1. Nature **542**, 456–460 (2017). https://doi.org/10.1038/nature21360
42. R.K. Kopparapu, R. Ramirez, J.F. Kasting et al., Habitable Zones around main-sequence stars: new estimates. ApJ **765**, 131 16, pp. (2013). https://iopscience.iop.org/article/10.1088/0004-637X/765/2/131/meta
43. M. Turbet, E. Bolmont, G. Chaverot et al., Day-night cloud asymmetry prevents early oceans on Venus but not on Earth. Nature **598**, 276–280 (2021). https://doi.org/10.1038/s41586-021-03873-w
44. D. Charbonneau, T.M. Brown, R.W. Noyes, R.L. Gilliland, The detection of an Extrasolar Planet Atmosphere. ApJ **568**, 377–384 (2002). https://doi.org/10.1086/338770
45. A. Vidal-Madjar, A. Lecavelier des Etangs, J.M. Désert, An extended upper atmosphere around the extrasolar planet HD209458b. Nature **422**, 143–146 (2003). https://doi.org/10.1038/nature01448
46. D.K. Sing, J.J. Fortney, N. Nikolov et al., A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. Nature **529**, 59–62 (2016). https://doi.org/10.1038/nature16068
47. S. Jin, C. Mordasini, Compositional imprints in density-distance-time: a rocky composition for close-in low-mass exoplanets from the location of the valley of evaporation. ApJ **853**, 163, 23 pp. (2018). https://doi.org/10.3847/1538-4357/aa9f1e
48. J.E. Owen, Y. Wu, The evaporation valley in the Kepler planets. ApJ **847**, 29. 14 pp. (2017). https://doi.org/10.3847/1538-4357/aa890a
49. I.A.G. Snellen, R.J. de Kok, E.J.W. de Mooij, S. Albrecht, The orbital motion, absolute mass and high-altitude winds of exoplanet HD209458b. Nature **465**, 1049–1051 (2010). https://doi.org/10.1038/nature09111
50. D. Ehrenreich, C. Lovis, R. Allart et al., Nightside condensation of iron in an ultrahot giant exoplanet. Nature **580**, 597–601 (2020). https://doi.org/10.1038/s41586-020-2107-1
51. H. Lammer, F. Selsis, I. Ribas et al., Atmospheric loss of exoplanets resulting from stellar x-ray and extreme-ultraviolet heating. ApJL **598**, L121–L124 (2003)
52. J.E. Owen, Atmospheric Escape and the Evolution of Close-In Exoplanets. Annu. Rev. Earth Planet. Sci. **47**, 67–90 (2019). https://doi.org/10.1146/annurev-earth-053018-060246
53. A. Lecavelier des Etangs, A. Vidal-Madjar, J.C. McConnel, G. Hébrard, Atmospheric escape from hot Jupiters. A&A L. **418**, L1–L4 (2004). ArXiv:astro.ph/0440.3369 v1
54. D. Ehrenreich, V. Bourrier, P. J. Wheatley et al., A giant comet-like cloud of hydrogen escaping the warm Neptune-mass exoplanet GJ 436b. Nature **522**, 459 (2015). https://doi.org/10.1038/nature14501
55. V. Bourrier, A. Lecavelier des Etangs, D. Ehrenreich et al., An evaporating planet in the wind: stellar wind interactions with the radiatively braked exosphere of GJ 436 b. A&A **591**, 121 pp. 14 (2016). https://doi.org/10.1051/0004/6361/201628362
56. R. Yelle, Aeronomy of extra-solar giant planets at small orbital distances. Icarus **170**, 167–179 (2004). https://www.sciencedirect.com/science/article/abs/pii/S0019103504000727
57. V. Bourrier, Lecavelier des Etangs, D. Ehrenreich et al, Hubble PanCET: An extended upper atmosphere of neutral hydrogen around the warm Neptune GJ 3470 b. A&A **620**, A147 (2018). https://doi.org/10.151/0004-6361/201833675

58. A. Vidal-Madjar, J.M. Désert, A. Lecavelier des Etangs, et al., Detection of oxygen and carbon in the hydrodynamically escaping atmosphere of the extrasolar planet HD 209458b. ApJ **604**, L69 (2004)
59. L. Ben-Jaffel, G.E. Ballester, Transit of exomoon plasma tori: new diagnosis. ApJL **785**, L30. 6pp. (2014). https://doi.org/10.1088/2041-8205/785/2/L30
60. L. Fossati, C.A. Haswell, C.S. Froing et al., Metals in the exosphere of the highly irradiated planet WASP-12b. ApJL **714**, L222–L224 (2010). https://doi.org/10.1088/2041-8205/714/2/L222
61. A. García Muñoz, P.C. Schneider, Rapid escape of ultra-hot exoplanet atmospheres driven by Hydrogen Balmer absorption. ApJ **884** L43, 13 pp. (2019). https://doi.org/10.3847/2041-8213/ab498d
62. M.L. Khodachenko, I.F. Shaikhislamov, Lammer, H et al. (2017) Lyα Absorption at Transits of HD 209458b: A Comparative Study of Various Mechanisms Under Different Conditions. ApJ vol. 847, 126. 13pp. https://doi.org/10.3847/1538-4357/aa88ad
63. A. Wyttenbach, D. Ehrenreich, C. Lovis et al., Spectrally resolved detection of sodium in the atmosphere of HD 189733b with the HARPS spectrograph. **A&A** 577, A62, 13 pp (2015). https://doi.org/10.1051/0004-6361/201525729
64. A. Oklopčić, C.M. Hirata, A New Window into Escaping Exoplanet Atmospheres: 10830 Å Line of Helium. ApJ L. **855**, L11. 7 pp. (2018). https://doi.org/10.3847/2041-8213/aaada9
65. J.J. Spake, D.K. Sing, T.M. Evans et al., Helium in the eroding atmosphere of an exoplanet. Nature **557**, 68–70 (2018). https://doi.org/10.1038/s41586-018-0067-5
66. R. Allart, V. Bourrier, C. Lovis et al., Spectral resolved helium absorption signature from the extended atmosphere of a warm Neptune-mass exoplanet. Science **362**(6421), 1384–1387 (2018). https://doi.org/10.1126/science.aat5879
67. L. Nortmann, E. Pallé, M. Salz et al., Ground-based detection of an extended helium atmosphere in the Saturn-mass exoplanet WASP-69b. Science **362**(6421), 1388–1391 (2018). https://doi.org/10.1126/science.aat5348
68. M. Salz, S. Czesla, P.C. Schneider et al., Detection of He I λ10830 Å absorption on HD 189733 b with CARMENES high-resolution transmission spectroscopy. A&A **620**, A97. 13 pp. (2018). https://doi.org/10.1051/0004-6361/201833694
69. F.J. Alonso-Floriano, I.A.G. Snellen, S. Czesla, et al., He I λ 10830 Å in the transmission spectrum of HD 209458 b. A&A **629**, A110 (2019). https://doi.org/10.105170004-6361/201935979. arXiv:1907.13425
70. M. Mayor, M. Marmier, C. Lovis et al., The HARPS search for southern extra-solar planets XXXIV, in *Occurrence, Mass Distribution and Orbital Properties of Super-Earths and Neptune-Mass Planets* (2011). arXiv:1109.2497
71. G.D. Mulders, C. Mordasini, I. Pascucci et al., The Exoplanet Population Observation Simulator. II. Population Synthesis in the Era of Kepler. ApJ **887**(2), 157, 14 pp. (2019). https://iopscience.iop.org/article/10.3847/1538-4357/ab5187/pdf
72. W. Zhu, S. Dong, Exoplanet statistics and theoretical implications. Ann. Rev. A&A **59**, 291–336 (2021). https://doi.org/10.1146/annurev-astro-112420-020055
73. K. Tsiganis, R. Gomes, A. Morbidelli, H.F. Levison, Origin of the orbital architecture of the giant planets of the solar system. Nature **435**, 459–461 (2005). https://doi.org/10.1038/nature03539
74. K.J. Walsh, A. Morbidelli, S.N. Raymond et al., A low mass for Mars from Jupiter's early gas-driven migration. Nature **475**, 206–209 (2011). https://doi.org/10.1038/nature10201
75. A.P. Boss, Proximity of Jupiter-Like Planets to Low Mass Stars. Science **267**, 360–362 (1995). https://doi.org/10.1126/science.267.5196.360
76. S.H. Dole, Computer simulation of the formation of planetary systems. Icarus **13**, 494–508 (1970). https://doi.org/10.1016/0019-1035(70)90095-3
77. C. Mordasini, H. Klahr, Y. Alibert, et al., Theory of planet formation. in Circumstellar Disks and Planets: Science Cases for the Second Generation VLTI Instrumentation, ed. by S. Wolf, Dec. 2010. arXiv:1012.5281.
78. F.H. Shu, Self-similar collapse of isothermal spheres and star formation. ApJ **214**, 488 (1977)

79. M.R. Bate, On the diversity and statistical properties of protostellar discs. Mon. Not. R. Astron. Soc. **475**(4), 5618–5658 (2018). https://doi.org/10.1093/mnras/sty169
80. C.J. Lada, Formation: from OB association to protostars, in *Proceedings Symposium of the International Astronomical Union*, vol. 115 (Star Forming Region, 1987), pp. 1–18. https://doi.org/10.1017/S0074180900094766
81. J.P. Williams, A. Cieza, Protoplanetary disks and their evolution. Ann. Rev. A&A **49**, 67–117 (2011). https://doi.org/10.1146/annurev-astro-081710-102548
82. S.M. Andrews, J. Huang, L.M. Perez et al., The disk substructures at high angular resolution project (DSHARP). I. Motivation, sample calibration, and overview. ApJL **869**, L41. 15 pp. (2018). https://doi.org/10.3847/2041-8213/aaf741
83. R. Lüst, Die Entwicklung einer um einen Zentralkörper rotierenden gasmasse. Z. Naturforschg **7A**, 87–98 (1952). DOI: https://doi.org&10515/zna-1952-0118
84. D. Lynden-Bell, J. E. Pringle, The Evolution of viscous discs and the origin of the Nebular variables, Mon. Nat. R. Astr. Soc. **168**, 603–637 (1974). DOI: https://doi.org/10.1093/mnras/168.3.603
85. C. Baruteau, A. Crida, S.J. Paardekooper et al., Planet-Disk Interactions and Early Evolution of Planetary Systems. Protostars and Planets VI, vol. 667, 24 pp. (Arizona University Press, Tucson, 2014). arXiv:1312.4293
86. T. Birnstiel, H. Klahr, B. Ercolano, A simple model for the evolution of the dust population in protoplanetary disks. A&A **539**, 148, 12 pp. (2012). https://doi.org/10.1051/0004-6361/201118136
87. H. Mizuno, Formation of the giant planets. Prog. Theoret. Phys. **64**(2), 544–557 (1980). https://doi.org/10.1143/PTP.64.544
88. F. Perri, A.G.W. Cameron, Hydrodynamic instability of the solar nebula in the presence of a planetary core. Icarus **22**(4), 416–425 (1974). https://doi.org/10.1016/0019-1035(74)90074-8
89. T. Birnstiel, C.P. Dullemond, F. Brauer, Gas-and dust evolution in protoplanetary disks. A&A **513**, 79. 21(2010). https://doi.org/10.1051/0004-6361/200913731
90. S. Weidenschilling, Aerodynamics of solid bodies in the solar nebula. Mon. Not. R. Astronom. Soc. **180**, 57–70 (1977a). https://doi.org/10.1093/MNRAS/180.2.57
91. L. Testi, T. Birnstiel, L. Ricci et al., *Dust Evolution in Protoplanetary Disks. Protostars and Planets VI*, ed. by H. Beuther, R.H. Klessen, C.P. Dullemond, T. Henning T, vol. 914, pp. 339–361 (2014)
92. A.N. Youdin, J. Goodman, Streaming instabilities in protoplanetary disks. ApJ **620**, 459–469 (2005). https://iopscience.iop.org/article/10.1086/426895
93. A. Johansen, J.S. Oishi, M.M. Low et al., Rapid planetesimal formation in turbulent circumstellar discs. Nature **448**, 1022–1025 (2007). https://doi.org/10.1038/nature06086
94. V.S. Safronov, Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets. (Keter Publishing House, 1972). ISBN:978–0–7065–1225–0
95. H. Tanaka, T. Takeuchi, W.R. Ward, Three-Dimensional Interaction between a Planet and an Isothermal Gaseous Disk. I. Corotation and Lindblad Torques and Planet Migration, ApJ **565**, 1257–1274 (2002). https://iopscience.iop.org/article/ and https://iopscience.iop.org/article/10.1086/324713
96. W.R. Ward, Survival of Planetary Systems. ApJ **482**, L211–L214 (1997). https://iopscience.iop.org/article/https://iopscience.iop.org/article/10.1086/310701/fulltext/5046.text.html/
97. F.S. Masset, A. Morbidelli, A. Crida, J. Ferreira, Disk Surface Density Transitios as Protoplanets Traps, ApJ **642**, 478–487 (2006)
98. S.J. Paardekooper, C. Baruteau, W. Cley, A torque formula for non-isothermal Type I planetary migration-II. Effects of diffusion. Mon. Not. R. Astron. Soc, **410**(1), 293–303 (2011). https://doi.org/10.1111/j.1365-2966.2010.17442.x
99. K.M. Dittkrist, C. Mordasini, H. Klahr et al., Impacts of planet migration models on planetary populations. Effects of saturation, cooling and stellar irradiation. A&A **567**, A121. 18 pp. (2014). https://doi.org/10.1051/0004-6361/201322506
100. M. Ogihara, A. Morbidelli, T. Guillot, Suppression of type I migration by disk winds. A&A **584**, L1 (2015). https://doi.org/10.1051/0004-636/201527117

101. J.L. Lissauer, Planet formation. Annu. Rev. Astr. Astrophy. **11**, 129–172 (1993)
102. E.W. Thommes, M.J. Duncan, H.F. Levison, Oligarchic growth of giant planets. Icarus **161**(2), 431–455 (2003). https://arxiv.org/pdf/astro-ph/0303269.pdf
103. C. Hayashi, Structure of the solar nebula, growth and decay of magnetic fields and effects of magnetic and turbulent viscosities on the nebula. Prog. Theor. Phys. Suppl. **70**, 35–53 (1981). https://doi.org/10.1143/PTPS.70.35
104. S. Weidenschilling, The distribution of mass in the planetary system and solar nebula. *Astrophysics and Space Science*, vol. 51, pp. 153–158. (Springer, 1977b). https://doi.org/10.1007/BF00642464
105. Y. Alibert, C. Mordasini, W. Benz, C. Winisdoerffer, Models of giant planet formation with migration and disc evolution. A&A **434**, 343–353 (2005). https://doi.org/10.1051/0004-6361:20042032
106. J.B. Pollack, O. Hubickyj, P. Bodenheimer et al., Formation of the giant planets by concurrent accretion of solids and gas. Icarus **124**, 62–85 (1996). https://doi.org/10.1006/icar.1996.0190
107. M. Ikoma, K. Nakazawa, H. Emori, Formation of giant planets: Dependences on core accretion rate and grain opacity. ApJ **537**, 1013–1025 (2000)
108. P.H. Bodenheimer, P. Pollack, Calculations of the accretion and evolution of giant planets the effects of solid cores. Icarus **67**, 391–408 (1986). https://doi.org/10.1016/0019-1035(86)90122-3
109. K.E. Haisch, E.A. Lada, C.J. Lada, Disk frequencies and lifetimes in young clusters. ApJ **553**, L153–L156 (2001)
110. W. Benz, W.L. Slattery, A.G.W. Cameron, The origin of the moon and the single-impact hypothesis I. Icarus **66**, 515–535 (1986). DOI: https://doi.org/10.1016/0019-1035(86)90088-6
111. J. Laskar, Large scale chaos and the spacing of the inner planets. Astron. Astrophys. **317**, L75–L78 (1997)
112. J. Laskar, Chaotic diffusion in the solar system. Icarus **196**(1), 1–15 (2008). https://doi.org/10.1016/j.icarus.2008.02.017
113. J.J. Fortney, M. Ikoma, Nettelmann et al., Self-consistent model atmospheres and the cooling of the solar system's giant planets. ApJ **729**, 32, 14 pp. (2011). https://doi.org/10.1088/0004-637X/729/1/32
114. D.J. Bower, D. Kitzmann, A.S. Wolf et al., Linking the evolution of terrestrial interiors and an early outgassed atmosphere to astrophysical observations. A&A **631**, A103, 18pp (2019). https://doi.org/10.1051/0004-6361/201935710
115. C.W. Ormel, H.H. Klahr, The effect of gas drag on the growth of protoplanets. Analytical expressions for the accretion of small bodies in laminar disks. A&A **520**, A43. 15 pp. (2010). https://doi.org/10.1051/0004-6361/201014903
116. A. Johansen, M. Lambrechts, Forming planets via pebble accretion. Annu. Rev. Earth Planet. Sci. **45**, 359–387 (2017). https://doi.org/10.1146/annurev-earth-063016-020226
117. G.A.L. Coleman, From dust to planets—I. Planetesimal and embryo formation. Mon. Not. Royal Astron. Soc. **506**, 3596–3614 (2021). DOI: https://doi.org/10.10937MNRAS/STAB1904
118. C. Marois, B. Macintosh, T. Barman et al., Direct imaging of multiple planets orbiting the star HR8799. Science **322**(5906), 1348–1352 (2008). https://doi.org/10.1126/science.1166585
119. A.P. Boss, Giant planet formation by gravitational instability. Science **276**(5320), 1836–1839 (1997). https://doi.org/10.1126/science.276.5320.1836
120. A.G.W. Cameron, Physics of the primitive solar accretion disk. Moon Planets 18, 5–40 (Springer, 1978). https://doi.org/10.1007/BF00896696
121. G.P. Kuiper, On the origin of the solar system. PNAS **37**(4), 1–14 (1951) and PNAS 37(4):233. https://doi.org/10.1073/pnas.37.1.1
122. L. Mayer, G. Lufkin, T. Quinn, J. Wadsley, Fragmentation of gravitational unstable gaseous protoplanetary disks with radiative transfer. AsJ **661**, L77–L80 (2007)
123. R. Helled, P. Bodenheimer, The effects of metallicity and grain growth and settling on the early evolution of gaseous protoplanets. Icarus **211**, 939–947 (2011). https://doi.org/10.1016/j.icarus.2010.09.024

124. O. Schib, C. Mordasini, N. Wenger et al., The influence of infall on the properties of protoplanetary discs. Statistics of masses, sizes, lifetimes, and fragmentation. A&A **645**, A43 (2021). https://doi.org/10.1051/0004-6361/202039154
125. K.M. Kratter, R.A. Murray-Clay, A.N. Youdin, The runts of the litter: why planets formed through gravitational instability can only be failed binary stars. ApJ **710**, 1375–1386 (2010)
126. S. Ida, D.N.C. Lin, Toward a deterministic model of planetary formation. II. the formation and retention of gas giant planets around stars with a range of metallicities. ApJ **616**, 567–572 (2004)
127. Y. Alibert, C. Mordasini , W. Benz, Migration and giant planet formation. A&A, **417**(1) L25–L28 (2004). DOI: https://doi.org/10.1051/0004-6361:20040053
128. A. Emsenhuber, C. Mordasini, R. Burn et al., The New Generation Planetary Population Synthesis (NGPPS). II. Planetary population of solar-like stars and overview of statistical results. A&A **656**, A70, 38 pp. (2021b). https://doi.org/10.1051/0004-6361/202038863
129. C. Mordasini, P. Molliere, K.M. Dittkrist et al., Global models of planet formation and evolution. Int. J. Astrobiol. **14**(2): Special Issue: Exoplanets, 201–233 (2015). Published online by Cambridge University Press. https://doi.org/10.1017/S1473550414000263
130. A. Emsenhuber, C. Mordasini, R. Burn et al., The new generation planetary population synthesis (NGPPS). I. Bern global model of planet formation and evolution, model tests, and emerging planetary systems. A&A, **656**, A69, 44pp (2021a). htpps://doi.org/https://doi.org/10.1051/0004-6361/202038553
131. Y. Alibert, F. Carron, A. Fortier et al., Theoretical models of planetary system formation: mass versus semi-major axis. A&A **558**, 109, 13 pp. (2013). https://doi.org/10.1051/0004-6361/201321690
132. C. Mordasini, Y. Alibert, W. Benz et al., Extrasolar planet population synthesis. II. Statistical comparison with observations. A&A **501**(3), 1161–1184 (2009). https://doi.org/10.1051/0004-6361/200810697
133. C. Mordasini, Planetary population synthesis. Handbook of exoplanets, 2425–2474, in Handbook of Exoplanets, ed. by H. Degg, J. Belmonte (Springer, Cham, 2018). https://doi.org/10.1007/978-3-319-55333-7_143 Print ISBN:978–3–319–55332–0 online ISBN: 978–3–319–55333–7 and arXiv:1804.01532v1
134. S.M. Andrews, D.J. Wilner, A.M. Hughes et al., Protoplanetary disk structures in ophiuchus. II. ApJ **723**, 1241–1254 (2010). https://doi.org/10.1088/0004-637X/723/2/1241
135. W. Benz, S. Ida, Y. Alibert et al., Planet population synthesis. *Protostars and Planets VI*, ed. by H. Beuther et al. (University of Arizona Press, Tucson, 2014), 691–713. https://doi.org/10.2458/azu_uapress_9780816531240-ch030. arXiv:1402.7086
136. O. Voelkel, H. Klahr, C. Mordasini, A. Emsenhuber, On the multiple Generations of Planetary Embryos. (2022). arXiv:2202.01500v1 https://doi.org/10.1088/0004-637X/723/2/1241. Exploring multiple generations of planetary embryos. published by EDP Science https://doi.org/10.105/0004-6361/202141830
137. N.J. Turner, S. Fromang, C. Gammie, Transport and accretion in planet-forming disks. Protostars and Planets VI, 411–432, ed. by H. Beuter et al. (University of Arizona, Tucson, 2014). https://doi.org/10.2458/azu_uapress_9780816531240-ch018
138. M. Keppler, M. Benisty, A. Müller et al., Discovery of a planetary-mass companion within the gap of the transition disk around PDS70. A&A, **617**, A44 (2018). DOI: https://doi.org/10.1051/0004-6361/201832957
139. J.F. Kasting, D.P. Whitmire, R.T. Reynolds, Habitable zones around main sequence stars. Icarus **101**(1), 108–128 (1993). https://doi.org/10.1006/icar.1993.1010
140. G. Anglada-Escudé, P.J. Amado, J. Barnes et al., A terrestrial planet candidate in a temperate orbit around Proxima Centauri. Nature **536**(437), 440 (2016). https://doi.org/10.1038/nature19106
141. V.S. Meadows, Reflections on O2 as a biosignature in exoplanetary atmospheres. *Astrobiology* vol. 17, n 10 (Mary Ann Liebert, Inc. Publishers, 2017). https://doi.org/10.1089/ast.2016.1578
142. C.H.L. Patty, J.G. Kühn, P.H. Lambrev et al., Biosignatures of the Earth. I. Airborne spectropolarimetric detection of photosynthetic life. A&A **651**, A68. 7pp. (2021). https://doi.org/10.1051/0004-6361/202140845

143. T.J. Fauchez, M. Turbet, G.L. Villanueva et al., Impact of clouds and hazes on the simulated JWST transmission spectra of habitable zone planets in the TRAPPIST-1 system. ApJ **887**, 194. 27 pp. (2019). https://doi.org/10.3847/1538-4357/ab5862
144. M. Turbet, E. Bolmont, V. Bourrier et al., A Review of Possible Planetary Atmospheres in the TRAPPIST-1 System. Space Science Reviews, **216**, 100 (2020). DOI: https://doi.org/10.1007/s11214-020-00719-1
145. I. Snellen, R. de Kok, J.L. Birkby et al., Combining high-dispersion spectroscopy with high contrast imaging: probing rocky planets around our nearest neighbors. A&A **576**, A59 (2015). https://doi.org/10.1051/0004-6361/201425018
146. C. Lovis, I. Snellen, D. Mouillet et al., Atmospheric characterization of Proxima b by coupling the SPHERE high-contrast imager to the ESPRESSO spectrograph. A&A **599**, A16, 16 pp. (2017). https://doi.org/10.1051/0004-6361/201629682
147. A. Segura, L.M. Walkowicz, V. Meadows et al., The effect of a strong stellar flare on the atmospheric chemistry of an earth-like planet orbiting an M Dwarf. Astrobiology **10**(7), 751–771 (2010). https://doi.org/10.1089/ast.2009.0376
148. S.P. Quanz, M. Ottiger, E. Fontanet et al., *Large Interferometer for Exoplanets (LIFE): I. Improved exoplanet detection yield estimates for a large mid-infrared space-interferometer mission* A&A 664, A21, 22pp (2022). https://doi.org/10.1051/0004-6361/202140366 . (Published online 09 August 2022)
149. A. Romero-Wolf, G. Bryden, S. Seager et al., Starshade rendezvous: exoplanet sensitivity and observing strategy. J. Astron. Telesc. Inst. Syst. **7**(2), 021210 (2021). https://doi.org/10.1117/1.JATIS.7.2.021210

Philosophy

Time, Space and Matter in the Primordial Universe

Francesca Vidotto

1 Introduction

General Relativity and Quantum Mechanics are the two great revolutions of XX century physics. Both have changed our perspective on the universe. Both have expanded the boundaries of the universe, one towards large scales, the other towards small scales. They are conceptual revolutions in the way we understand space and time, and matter, respectively. Shaking the very foundations of physics, these two revolutions have opened the possibility of promoting our knowledge of the universe as a whole to an exact science. The birth of modern cosmology is inextricably linked to what we have understood in relativity and quantum theory.

A remarkable aspect of modern cosmology is precisely that it needs to combine our knowledge about small scales and our knowledge about large scales. Only this convergence can give us a coherent vision of the universe. Convergence and coherence reinforce our confidence in the physical theories we use. In this sense, cosmology is a science that brings everything together: all the physics we know.

This does not mean that cosmology is the science that describes everything, namely everything the universe contains [1]. On the contrary, cosmology was born understanding that it is possible to describe the universe at its largest accessible scale with simple laws, by treating it as a whole and leaving out the details. The embodiment of this idea is Einstein's "cosmological principle" [2]. This principle allows to apply Einstein's theory of gravitation to the whole universe. This is an application of General Relativity where we can use exact solutions. The idea is simple and ingenious: it is possible to capture the dynamics of the universe using a single quantity,

Western University is located in the traditional lands of Anishinaabek, Haudenosaunee, Lūunaapéewak and Attawandaron peoples.

F. Vidotto (✉)
Department of Physics and Astronomy, Department of Philosophy, Rotman Institute of Philosophy, The University of Western Ontario, London, Canada
e-mail: fvidotto@uwo.ca

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_18

for example the radius or the volume. More precisely, it is possible to reduce the infinite number of degrees of freedom of the gravitational field down to a model with a single degree of freedom. This is enough to capture the large scale dynamics of the universe, that is, how the universe evolves as a whole.

This is what Einstein discovered in 1917: it is possible to write a single equation to describe the large scale dynamics of the universe. This equation shows that the universe must evolve, but Einstein resisted what his own equations were indicating, and it was only later, thanks to Georges Lemaître [3], that it was realized that indeed the universe could not be static but had to expand. On the other hand this corresponds to the heart of Einstein's General Relativity: space-time is identified with the gravitational field [4], and this, like all fields, has a dynamic nature.

The gravitational field depends on the presence of matter. The physical approximation that Einstein takes, with the cosmological principle, consists in assuming that, on large scale, matter can be considered as uniformly distributed. Einstein imagines that the matter uniformly distributed are stars that fill the whole universe. Today we know that Einstein's assumption is valid, but for the very large scale distribution of galaxies, not stars. In fact, it is only almost 10 years after Einstein writes his equations that it is understood that nebulae are other galaxies [5]: the universe is far bigger than we thought. The step taken by Einstein introducing the cosmological principle can be interpreted in different ways. It is an approximation, which is not valid within our galaxy or with respect to the local group of galaxies of which the Milky Way is a part. But it is an approximation that works well on the largest scale, where the distribution of matter we consider is given by the average distribution of galaxy clusters.

It is largely a philosophical motivation that drives Einstein to take the cosmological principle as the basis of his cosmology [6]. The cosmological principle tells us that there is no privileged observer in the universe, hence all observers see the same universe. This leads to a homogeneous and isotropic universe. This is a reinforced version of the Copernican principle, according to which our position on Earth is not a privileged one. Einstein is guided by the philosophical ideas of Ernst Mach [7]. According to Mach, the inertia of bodies should be accounted by the distribution of matter in the universe. Einstein had arrived at the equations of General Relativity and the idea that physics should not depend on the choice of coordinates, under the influence of the ideas of Mach. He generalizes the idea that there are no privileged points in the universe, to the idea that there are no privileged reference systems of the universe. This way of looking at things proves extremely fruitful. It leads to a simplification of our vision of the world in which there are only general covariant substances: such must be the nature at the fundamental level of quantum fields as well as of the universe itself.

In the rest of the article, I discuss this understanding about time, space, and matter, based on what we know from General Relativity and quantum theory, particularly how this affects our understanding of the universe.

2 Time

The central idea of General Relativity is the identification of space-time with the gravitational field. Distances (that in non relativistic physics are given by the cartesian coordinates x, y, z) and durations (that in non relativistic physics are given by the cartesian coordinate t) depend on the gravitational field. Hence measurable physical time is a dynamical variable, on the same level as the other physical variables. This fact is one of the most important and disconcerting aspects of General Relativity. Special Relativity already relates space and time, conceptually joining them into a single entity, so that it is no longer possible to think of time in absolute terms as in Newtonian mechanics. But in General Relativity there is a far more radical step: frames of reference are local, they cannot be extended to the whole universe, and can only be valid in a limited region of space time. In particular, we have to associate a different time with every local frame of reference: time becomes multi-fingered. Instead of viewing it as a single abstract quantity, we have to consider it as the local reading of physical quantities that we take as *clocks*.

From a formal point of view, what happens in General Relativity, is that in the Hamiltonian formalism the evolution is coded in a set of constraints [8]. The coordinate time (as opposed to the physical time measured by clocks) is a dispensable gauge dependent quantity. It is often said in General Relativity or in quantum gravity that time disappears from the equations. This does not mean that these theories do not describe evolution; it means that the theory describes the evolution with respect to a dynamic variable that is identified as the clock. There is nothing mysterious about this, yet the disappearance of time has confused many scientists, also in the development of cosmological models. In cosmology, it is possible to choose a privileged global clock by exploiting the homogeneity and isotropy of the universe on a large scale. This is a consequence of the specific approximations made in classical cosmology. In this way in cosmology it is possible to choose a preferred time with respect to which to write the evolution equations. But this ceases to be true when we promote these same equations to quantum equations: it is no longer possible to choose the privileged time of cosmology and we must return to a writing of the equations in terms of constraints [9].

In the early universe, quantum effects on the dynamics of the gravitational field cannot be disregarded. To describe them we need the quantum version of the equations of the gravitational field. We know that all physical fields are quantum fields, but while we have equations that describe the quantum behavior of the fields of matter well, the quantum equations of the gravitational field have not yet found a definitive form. The most famous equation of quantum gravity is the Wheeler-deWitt equation [10]. When it was written, this equation puzzled physicists because the evolution it describes does not select any privileged time, so that this equation has become the prototype of a timeless equation. Half a century later, we now understand that it describes evolution is with respect to any dynamic variable, a clock, which is itself quantum.

The Wheeler-deWitt equation was not well defined and it is not easy to handle. It was initially studied in the simplified context of cosmology, where there is only one degree of freedom to evolve. The object that a physicist faces at this point is breathtaking: the equation describes a quantum universe evolving in quantum time. What may this mean?

A natural objection to a quantum universe is the idea that Quantum Mechanics applies only to small scales, and not to huge systems such as the universe. But this is misleading for two reasons. First, Quantum Mechanics does not apply only to small objects, quantum effects depend on size, energy and other variables. Second, the universe is expanding and it was very small in the early stages of this expansion. In fact, if we extrapolate backwards the history of the universe, we need to imagine that all the matter in the universe was contained in the size of a proton. This implies an energy density for which there is no doubt that the universe behaves quantumly. This conclusion forces us to seriously consider the realm of quantum cosmology.

A characteristic phenomenon of quantum physics is superposition. When this is applied to macroscopic objects, quantum superposition generates the conceptual confusions illustrated by Schrödinger's famous cat which is in a superposition of two macroscopically different states. That is, its classical state is not defined. The way to clarify this confusion is still debated and affects our way of thinking of the early universe. In the case of a superimposed universe, we must think that the early universe was such that its geometry and its physical time were not defined. The fact that the universe is expanding today is encoded in its geometry; the time reversed geometry would be a universe that is contracting instead. In the early universe these two geometries may have been in a superposition. In the same way, it is possible for temporal directions to be in superposition. This makes it possible to switch from one geometry to another, from one time to another. This is a typical effect of Quantum Mechanics: tunnelling. From the point of view of the evolution of the universe, this allows for a universe that first contracts, then bounces and begins to expand [11].

The effects of quantum superposition on the structure of time are not confined to cosmology. Even in elementary non relativistic Quantum Mechanics it is possible to have the quantum superposition of two histories in which the temporal order in which to events happen is different. These are called indefinite causal structures [12]. In this sense "quantum time" is much closer to us than it might seem if confined to the early universe.

In cosmology, this superposition of causal structures generates another sense for which time disappears in quantum cosmology, distinct from the general relativistic multiplicity of clock times. If a clock time does not determine a unique causal structure, it looses its specificity as a temporal variable. Temporal and spacial directions get mixed up. In a universe with four dimensions all four of these dimensions are then on the same footing and the resulting geometry is no longer that of the space-time considered by Einstein (the space of Minkowski and Lorentz) but rather a space closer to Euclidean geometry in four dimensions. In a celebrated work [13], Stephen Hawking and James Hartle suggested that the early universe could have emerged form such an "Euclidean" phase, where the spacial and temporal directions of spacetime are no longer distinguished. Their suggestion for the early quantum

universe is known as "no-boundary" proposal. In this model, if we go back to the early universe, we reach a point where it is no longer possible to go further back because time disappears. This is the picture that Hawking illustrated in his famous book "A Brief History of Time".

The issue, however, is open, and much debated. It is possible that the "no-boundary" proposal captures an aspect of time in the early universe: the lack of a specific time structure. But it might also be that what this proposal captures only a piece of a story in which the universe transitions from a contracting phase to an expanding one. Also in this case there is "no-boundary", that is there is no beginning of time with which one collides by going further and further back in time. But in this case the universe may have had its own evolution before the Euclidean phase considered by Hartle and Hawking.

A further clarification is important. In the discussion so far, we have discussed causal structure. This is encoded in the metric, the mathematical object that contains the information on the gravitational field and on the geometry. When in the early universe geometry is transformed into Euclidean geometry, this is equivalent to a change in the signature of the metric. But caution: this causal structure captures the distinction between space and time direction in spacetime, it does not capture the *direction* of time. The direction of time is a phenomenon that is not described by General Relativity. Fields, including the gravitational one, are described by mechanical laws, be they classical or quantum. Mechanics, including Quantum Mechanics, is invariant under CPT, namely inversion of charge, parity and time inversions. Hence the dynamics does not have a privileged direction of time. Instead, the directionality of time is a characteristic of thermodynamic phenomenon. In the absence of heat sources, a hot body becomes cold; a very orderly system becomes disordered over time; a system that can be described with one or a few degrees of freedom will evolve into a system that requires many degrees of freedom to be described: in a single concept, observed entropy is always lower in the past and higher in the future. It is this thermodynamic concept of entropy that contains information on the direction of time.

In our universe we observe entropy to uniformly increase. Whatever the entropy of the universe today, this entropy must have been much lower in the past. This fact is known and discussed under the name of "Past Hypothesis" [14]. Since low entropy can be thought as a "special" condition, this fact raises the question why was the universe in special initial conditions in the past? The discussion is open and involves physicists and philosophers in equal measure. We do not have an answer yet, but my feeling is that we must continue on the path that General Relativity and Quantum Mechanics have indicated to us up to now: we must not think in absolute terms, but in relational terms. Perhaps this is teaching us, once again that we should not think of the conditions of the universe as special at first, but that this is an effect of our point of view and our particular way of interacting with matter and space-time.

3 Space

Einstein's relativity links space and time together, and space-time to matter. This link plays a fundamental role in the early universe for the formation of structures. The universe is homogeneous on a very large scale but it is not at other scales: gravity generates an instability in the homogeneity: any excess of matter is a source of gravitational attraction and becomes a seed for the formation of structures such as stars, black holes, galaxies, galaxy clusters, and so on. In the early universe there must therefore have been small deviations from global homogeneity. These inhomogeneities, which we can think of as deviations from a homogeneous distribution of matter, are now understood to be of quantum origin. When the universe comes out of its primordial quantum phase and becomes classical, the quantum fluctuations of matter turn into statistical fluctuations in the distribution of matter. The fact that matter has a quantum nature appears therefore to be an essential element in the explanation of the structures that populate the universe.

In Einstein's equations, space-time and matter are interconnected. The dynamics of space-time is determined by the presence of matter, and vice versa, matter moves according to the shape of space-time. Consequently, if matter has quantum properties, this link implies that geometry must have them too. What does it mean for spacetime to be quantized? Although to fully answer this question it is necessary to understand the quantization of whole space-time, it is easier to understand and explain it by focusing on space: let's fix an instant in time, and consider quantized space.

A characteristic phenomenon of quantum physics is the possibility for a system to be in a superposition of different states. This introduces a form of fuzziness or indeterminacy, into all quantum systems: physical variable can fail to have a determined value. We have seen how this can concretely manifest in the case of the geometry of space-time: in a great rebound, the universe can be in a state where it is simultaneously contracting and expanding. The superimposition of geometries, which plays a role in the description of the universe at very large scale, as well as on the scale of possible laboratory experiments too, becomes more and more important as we approach shorter scales, until it touches the smallest scale of all, that of Planck [15].

This leads us to another aspect of quantization, the most fundamental, together with superposition, which gives the theory its name: everything manifests itself in the form of discrete quantities, or "quanta". Let us remember the case of the electromagnetic field: the continuous field can be thought of as emerging from a set of photons, the quanta of the electromagnetic field. Similarly, in gravity, space-time can be thought of as emerging from a set of quanta of the gravitational field. It is necessary to be cautious in the way in which we speak of emergency in this context: both in the electromagnetic case and in the gravitational case, the emergence has the clear meaning of a classical limit. In other words, this is the familiar process by which we loose properties characteristic of the quantum regime, which are replaced by other characteristics of the classical regime. In particular, in the case of fields, we pass from a discrete description to a continuous one.

Can we say that space disappears in the early universe when the gravitational field enters a quantum regime? If we identify space with its property of continuity, then yes, it disappears. But the continuity property is not an essential property for the gravitational field. The continuity property is useful for describing contiguity and locality of processes at scales much larger than the discreteness scale. What remains of continuity and locality, in a quantized space? The information at the basis of continuity and locality must be rooted in some properties of the quantum states of space. In a sense, this is what guides us in their definition. The quantum states of the gravitational field, at a given time, must encode the information on which quantum of space is adjacent to which other, and which quantum is interacting with each other. In the case of Loop Quantum Gravity [16, 17] this idea is embodied in quantum networks (spinnetworks) which encode the contiguity relationships between the different quanta of space. The interactions between the quanta of space constitute a network which, in the classical limit, results in the familiar space described by General Relativity.

To what extent does each quantum space contain properties that correspond to those we attribute to classical space? Single quanta of space embody a notion of geometry (based on Roger Penrose's spin-geometry theorem [18]) and therefore a notion of distance, the key element in building the theory of General Relativity. Spin networks can thus be associated to a metric, but it will be a fuzzy metric. A quantum of space can also have information on the number of dimensions and on the type of geometry (Lorentzian or Euclidean). Just as a photon is already a manifestation of the electromagnetic field even when the classical limit of the theory is not considered, in the same way a quantum of space is in many respects a manifestation of the gravitational field, even when its classical regime is not considered.

An intuitive way of thinking about the quanta of the gravitational field is that each corresponds to a small unit of space: let's take a certain region of space and divide it into many tiny pieces, but not smaller than the Planck scale. The Planck area and the Planck volume, represent in fact a minimal size for space. This is analogous to the fact that the energy of photons of a given frequency is proportional to the Planck constant. But, just as for photons we cannot simply think of the electromagnetic field just as a classical sum of small Planck energy packets, in the same way we cannot think that space is simply constituted by adding up many Planck volumes.

The quantization that manifests itself for the quanta of space is genuine, it is not an artifice of having taken a discretization of space. This means for example that for each individual quantum of space, if we measure its properties such as its area or its volume, we have the characteristic quantum leaps between one value and another: For example, between the zero value of the volume and the first smaller one measured value, i.e. a jump, and between the first and the second there is another jump.

Let me illustrate this fact with a specific example. In Quantum Mechanics the angular momentum is quantized: it can only take certain values, in jumps. Notice that the fact that the components of the angular momentum are quantized does not break the symmetry of the theory under rotations. In a rotated frame, the eigenstates are rotated, but the eigenvalues are the same. Similarly, the fact that space is quantized does not break the characteristic symmetries of space and space-time—and in

particular the Lorenz symmetry is preserved [19]: when I observe the quanta of space in a moving frame of reference, I still find the same jumps quantum in the area and in the volume, but with different probabilities.

In Quantum Mechanics, as energy rises, the intervals between high energy states, namely the quantum leaps at higher energy, become smaller and less visible. Similarly in gravity, as a larger space is considered, the more quantum leaps in area and volume become smaller and less visible. When we consider macroscopic regions of space, and even more so considering cosmological scales, the granularity of space is negligible. The quanta of space have a characteristic scale which is the Planck scale, a very minute scale: there is the same distance in terms of orders of magnitude between us and the scale of a proton, as between the scale of the proton and that of a quantum of space.

And yet, the granularity of space plays an important role in the early universe. When we study the quantum fluctuations from which every structure derives, these fluctuations have to do not only with matter but through Einstein's equation also with geometry. To understand the quantum fluctuations of geometry and to study how they evolve it is necessary to give a description that takes into account the granularity of space [20, 21].

4 Matter

In the early universe, the essential feature of matter fields, namely the fact that they are quantized, needs to be extended to the gravitational field, which is also quantized. Conversely, from General Relativity, we have learnt another important property that a field must possess: it must not depend on the coordinates with which it is described. One way to express this is the fact that the theory is independent of background. The gravitational field provides a background to the fields of matter by choosing a fixed configuration, that is, by neglecting their dynamic nature. The presence of a background and the fact that this is continuous form the basis of the whole construction of the theory of fields, both classical and quantum. Describing matter at a fundamental level therefore means doing it in a way that respects general covariance and is compatible with the quantum discretion of space-time.

One way to achieve this is to think of spinnetwork states as lattices, albeit special lattices as they are invariant under diffeomorphisms, i.e. the coordinate transformations that characterize General Relativity. Precisely from the fundamental physics developed using lattices come some techniques at the basis of the quantization of gravity. One such method are Wilson loops, which are at the basis of Loop Quantum Gravity [22]. Wilson lines and loops are known objects in fundamental physics, especially for the study of strong interactions.

An interesting observation is that the loop quantization can be applied not only in the path-integral formalism but also in the standard Hamiltonian formulation of Quantum Mechanics, à la Dirac. In addition to the case of space quanta, we can use this same quantization technique for ordinary particles as well. It can be shown

that the usual quantization is found within the loop quantization when the loops become very short, that is, in a regime where their length can be neglected. The usual quantization simply becomes then a special case of the general case of loop quantization.

The quantization of space-time in Loop Quantum Gravity is compatible with the Standard Model of particles and with its possible extensions. In this sense, quantum gravity presents no problems and provides a coherent picture in which all fundamental fields are quantized. On the other hand, this provides no explanation for the presence of other fields that are hypothesized in cosmology to explain observations and to make the different cosmological models work. For example, quantum gravity does not give any explanation for what could be the field that gives the accelerated phase hypothesized in the early universe, the *inflaton* [23].

Some cosmological models postulate the existence of exotic matter, namely matter with equations for the energy that do not follow the ordinary ones. This serves to avoid the presence at the beginning of the expansion of the universe of a singularity, a point of infinite energy density that appears in the classical theory of gravity [24]. This is not necessary considering the quantization of gravity: space-time quanta implies that such singularities cannot exist. In a simple analogy, we can think that just as the electrons of an atom do not fall into its nucleus thanks to the quantization of energy, in the same way a gravitational collapse will not give rise to a singularity thanks to the quantization of space-time.

Note that in this presentation insofar dark energy was not mentioned. In fact, it is not necessary to think of dark energy as a form of matter. The reason is that the expansion of the universe, as far as we know today, is compatible with the cosmological constant that appears in the equations of General Relativity [25]. It is therefore an effect due to geometry, not to matter. It can be included in the quantum equations of gravity (in the context of Loop Quantum Gravity, this is not true in other contexts such as in the case of String Theories, where instead a positive cosmological constant is problematic).

A different consideration applies in the case of dark matter. This is, indeed, a big open question: Which matter only interacts gravitationally and yet constitutes most of the matter in our universe? Much research has been done to look for new particles that could explain it: unfortunately, to date, none of the particles hypothesized for this purpose have been found. An explanation in terms of some new form of matter becomes less and less credible. The alternative is to pay attention back to gravity. One possibility is to study modifications to Einstein's theory that explain all the observations on a galactic and cosmological scale. There are more or less successful attempts in this direction, even though the more we collect data to test General Relativity, the stronger the case insofar for no modifications.

One possibility, today among the favoured ones by physicists and which is simply based on a gravitational mechanism, is that dark matter may be simply black holes that formed in the primordial universe. This is today a very promising possibility, which could develop in different directions depending on what the mass (or mass distribution) of these primordial black holes is. There are several scenarios open today: both that of very large black holes, and that of small black holes. Unfortunately,

making a direct detection of these black holes is very difficult: they are scattered and elusive! Their credibility as dark matter is therefore closely linked to the viability of the cosmological models in which they can be produced.

Finally, one possibility that I find very exciting is that dark matter could be composed not of black holes, but of remnants [26] left over after the “death of black holes." This scenario is particularly interesting when compared with cosmological models where there is a bounce. The remnants, in this case, may have formed in the contraction phase of the universe and may have survived through the bouncing phase [27]. How a black hole exhausts its black hole life and turns into a remnant depends on a quantum phenomenon of geometry. In this sense, black hole remnants are effectively a dark matter candidate that requires quantum gravity to be fully explained. The beauty of this model, I find, is that it requires no changes to physics other than what we know and what we expect (we expect gravity to be quantized).

5 Conclusions

To understand time, space, and matter in the primordial universe, we must give up relying solely on our common sense. This is what great science does: to show us beyond the veil, beyond the patterns that are useful to us for daily life but which may not be effective in capturing the essence of natural phenomena. Understanding the quantum characteristics of the space and time is requiring to build a new theory, a theory of quantum gravity.

To build a new theory we have to rely on philosophy. In fact, it is this that gives us conceptual tools to discern what, in the theories currently available to us, is essential, and what is instead an accident of the particular situation in which we study a given phenomenon. We must therefore proceed by studying the fundamentals, asking ourselves what concepts of space and time offer their most essential version. What concepts can we do without to describe space and time? Which ones are indispensable? It is through this path of conceptual cleaning that we put ourselves in the position of writing the new equations in a language compatible with both Quantum Mechanics and General Relativity.

Because of this need to rethink the basis, the recent research in cosmology and quantum gravity has given rise to an intense interactions between physicists and philosophers of science. On the one hand, philosophers of science study tentative theories with attention to their conceptual coherence and their incompleteness. On the other hand, physicists find inspiration not only in these analysis, but also in the traditional philosophical discussion on the nature of spacetime and matter. The classical philosophy, also in part inspired by science, has developed a variety of different manners for interpreting the notions of space and time. From Aristotle to Descartes, from Hume to Kant, and from Mach to Reichenbach, philosophers have provided a variety of different conceptualizations of the temporal and spacial structure of the world. Familiarity with this variety has become an indispensable tool for the advancement of cosmology and fundamental physics. For instance the notion

of space as an entity, or as relations, or as an a priori condition for intuition, or as a tool to organize sensations or empirical data: all can be distinguished and each play a different role in a modern theory of gravity. The common sense and simple minded notion of space gets opened up into a complex set of concepts. Even richer is the construction of time, whose different properties are understood by different part of our science: from fundamental timeless quantum gravity to the general relativistic and Newtonian formulation, all the way down to thermodynamics and possibly even neurosciences.

Viceversa, this conceptual analysis is nourished and put to work by the scientific method, by observations and experiments. We certainly do not yet have direct experiments proving the quantum properties of space-time (although, as I mentioned, these are now closer than previously imagined). This does not mean that we do not have any experimental data: we have a series of data that constraint the space of possible theories of quantum gravity. For example, we know that no micro black holes are produced at CERN. We also know that, if they ever exist, supersymmetric particles do not manifest themselves at the energies available at CERN or through astrophysical phenomena associated with dark matter. We know with great accuracy that the Lorentz symmetry is not violated. We know that gravitational waves travel at the speed of light. And so on: Nature is giving us clues to follow.

Understanding space, time and matter in the early universe brings us to the edge of our knowledge: and it is there that philosophy and experience must meet.

References

1. F. Vidotto, Relational quantum cosmology, in *The Philosophy of Cosmology*, ed. by K. Chamcham, J. Silk, J.D. Barrow, S. Saunders (Cambridge University Press, 2017)
2. A. Einstein, *Kosmologische Betrachtungen Zur Allgemeinen Relativitätstheorie* (Preussische Akademie der Wissenschaften, Sitzungsberichte, 1917)
3. G. Lemaître, Un Univers Homogène de Masse Constante et de Rayon Croissant Rendant Compte de la Vitesse Radiale des Nébuleuses Extra-Galactiques *Annales de la Société scientifique de Bruxelles*, vol. A47, pp. 49–59
4. A. Einstein, Appendix V. Added to A. Einstein, *Relativity, the Special and the General Theory*, pp. 135–157 (1952)
5. E. Hubble, Cepheids in spiral nebulae. Observatory **48**, 139–142 (1925)
6. C. Smeenk, Einstein?s role in the creation of relativistic cosmology, In *Cambridge Companion to Einstein*, ed. by M. Janssen, C. Lehner (Cambridge University Press, 2012)
7. G. Mach, Einstein, and the search for reality. Daedalus **97**, 636–673 (1968)
8. C. Rovelli, *General Relativity: The Essentials* (Cambridge University Press, 2021)
9. C. Rovelli, The strange equation of quantum gravity. Class. Quant. Grav. **32**(12), 124005 (2015)
10. B. DeWitt, Quantum theory of gravity. I. The canonical theory. Phys. Rev. **160**(5), 1113–1148 (1967)
11. A. Ashtekar, An introduction to loop quantum gravity through cosmology. Nuovo Cim. B **122**, 135–155 (2007)
12. E. Castro-Ruiz, F. Giacomini, Časlav Brukner Dynamics of quantum causal structures. Phys. Rev. X **8**(1), 011047 (2018)
13. J. Hartle, S. Hawking, Wave Function of the Universe. Phys. Rev. D **28**, 2960–2975 (1983)
14. D. Albert, *Time and Chance* (Harvard University Press, 2000)

15. C. Rovelli, F. Vidotto, *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory* (Cambridge University Press, 2014)
16. R. Gambini, J. Pullin, *Loop Quantum Gravity for Everyone* (Oxford University Press, 2020)
17. R. Gambini, J. Pullin, *A First Course in Loop Quantum Gravity* (Oxford University Press, 2011)
18. R. Penrose, Combinatorial quantum theory and quantized directions, in *Advances in Twistor Theory*, ed. by L.P. Houghston, R.S. Ward (1966)
19. C. Rovelli, S. Speziale, Lorentz covariance of loop quantum gravity. Phys. Rev. D **83**, 104029 (2011)
20. A. Ashtekar, A. Barrau, Loop quantum cosmology: from pre-inflationary dynamics to observations. Class. Quant. Grav. **32**(23), 234001 (2015)
21. F. Gozzini, F. Vidotto, Primordial fluctuations from quantum gravity. Front. Astron. Space Sci. **7**, 629466 (2021)
22. C. Rovelli, *Quantum Gravity* (Cambridge University Press, 2004)
23. A. Guth, The inflationary universe: a possible solution to the horizon and flatness problems. Phys. Rev. D **23**, 347–356 (1981)
24. F. Tipler, Energy conditions and spacetime singularities. Phys. Rev. D **17**, 2521 (1978)
25. E. Bianchi, C. Rovelli, Why all these prejudices against a constant? (2010). arXiv:1002.3966
26. C. Rovelli, F. Vidotto, Small black/white hole stability and dark matter. Universe **4** (11), 127 (2018)
27. C. Rovelli, F. Vidotto, Pre-Big-Bang black-hole remnants and past low entropy. Universe **4** (11), 129 (2018)

Art

The Shore Between Art and Science

Enrico Magnani

The most beautiful experience we can have is the mysterious. It is the fundamental emotion that stands at the cradle of true art and true science. He who knows it not and can no longer wonder, no longer feel amazement, is as good as dead, a snuffed out candle. (Albert Einstein) [1]

Art and science are two aspects of human knowledge based on continuous research. Scientific ideas, discoveries and technologies nurtured my mind firstly as a researcher and now as an artist. Curiosity and intuition are precious allies and are accompanied by the search for beauty and the need to discover the unknown. The common traits in artistic and scientific research as well as the synergy between art and science will be followed by the description on how, as an artist, I describe some of today's challenges in cosmology. The presentation of the artworks inspired by supernovae, dark matter, multiverse and the project 'Quintessence', on permanent display at the Gran Sasso National Laboratory (LNGS), will depict my artistic research.

1 Introduction

For[1] a long time the world of art and the world of science have been separated and there was a perceived need to build a bridge between them. Therefore, initially, for my essay, I had thought of the title 'A bridge between art and science'. On reflection, however, as the integration and the communication between these two worlds is already undergoing—we just need to pay attention—I thought a more appropriate

[1] Artist website https://www.enricomagnani-art.com

E. Magnani (✉)
Reggio Emilia, Italy
e-mail: mail@enricomagnani-art.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Streit-Bianchi et al. (eds.), *Advances in Cosmology*,
https://doi.org/10.1007/978-3-031-05625-3_19

title would be 'The shore between art and science'. Indeed, a bridge is something that joins two separate banks that are distant, or difficult to reach, separated by a river or the sea. Art and science are instead two domains of human knowledge that do not overlap, but interpenetrate, just as happens on the shore when water and sand meet: they have imprecise and mobile boundaries. On one side of the shore there is sand, only sand; on the other side there is water, only water. But, on the shore, water and sand touch each other, mix together and get intertwined.

That the cosmos is a gigantic work of art is undeniable. It is not difficult to think of art when looking at the wonderful images collected by the Hubble telescope. Talking about the link between art and cosmology will be, however, like looking at the tip of the iceberg. In fact, in all domains of science—biology, chemistry, physics, mathematics, etc.—we can find aspects related to beauty, and not only the aesthetic beauty of images. When Paul Dirac, in his writings [2], talks about the 'principle of beauty'—as opposed to the 'principle of simplicity'—to be used to choose the equations of a possible physical model, he is talking about beauty related to mathematical aesthetics, which is something difficult to express in words, although those who are familiar with the profession have no difficulty in understanding what he is talking about. The aesthetics of images, on the other hand, is much more approachable by a heterogeneous audience although the perception of it may be subjective and time dependent. It is much easier for the public to appreciate an image, and even comment on it, than an equation. Perhaps this is because the equation presupposes studies of a certain kind, whereas the image, by its nature, is more immediate. Even the image, however, presupposes a preparatory study and knowledge to understand the representation, but often we let ourselves be carried away simply by the emotion generated by what we see, and sometimes that is enough to our eyes, our mind and our soul.

Let's take general relativity. By many accounts, it is the most beautiful theory ever conceived—the undisputed credit of one man: Albert Einstein. The equations of general relativity, for those who know mathematics and physics, have those aesthetic traits mentioned by Dirac in his writings. Unfortunately, it is difficult to really appreciate its beauty if you do not have scientific knowledge and sometimes even then it is not enough. Here, art can come into play acting as an ambassador of a message. Art can represent some ideas of general relativity and soften them, making them more accessible to the general public. The graphic representation of a universe curved and deformed by mass—although three-dimensional and not four-dimensional, as the equations tell us—is already a powerful foothold for a visualisation of the theory. This hold, thus, can lead the viewer to appreciate the scientific content even more, if not to instantly grasp its essence.

Today's cosmology and physics investigate complex and fascinating topics. Complexity is increasing, as is the need to make this complexity accessible. The artistic representation of scientific topics is in great development and the combination of art and science is giving rise to a period that many are calling a 'new renaissance'.

Dark matter and dark energy, multiple quantum or inflationary universes, are no longer just described in science fiction novels—and here I already wonder how

science fiction could have anticipated science. Cosmological theories, today, lay the foundations awaiting for their confirmation. The mathematical models describing them can appear really scary and for this reason keep the general public away from a world that instead deserves to be known, appreciated and understood. The artist has a task: make all this part of reality fascinating and colourful, accessible and, as physicists would say, 'sweeten the gradient' of access to this knowledge.

Often, however, the artist does not have adequate scientific training to represent science; in this case, it would be necessary to establish a real and more widespread dialogue as the one now occurring in big science centres. This is not surprising: as these two worlds have a lot in common. Not only the aesthetics, but also the very fact of investigating and trying to express the unknown and the mystery. Research, quest for knowledge and the use or development of tools and technologies is something that art and science certainly have in common although not at the same level of material resources invested, but certainly at a level of personal investment. Both, true art and true science, have the mission to expand the boundary of human knowledge, to move this boundary more and more towards the exploration of the unknown, thus transforming the unknown into the known.

Science is always on the quest to explain how the world works and what it is like, and deals with that, which is related to matter. Art, while also being linked to matter, can also afford to deal with what transcends it. It is impossible to think of an equation for love—at least for now—nevertheless love exists no less than planets or stars. A scientist must look at the artist as a researcher colleague. And the research of one can be of help to the research of the other. Let's think about cubism: when quantum mechanics made its appearance in the world, simultaneously, cubism, a new way of representing reality, appeared. One of the most characteristic ideas of the new physics was the wave-particle duality. An electron, for example, could be studied and seen both as a wave and as a particle. In art, with cubism, objects and people could be depicted from multiple perspectives; a face could be painted simultaneously in front and in profile.

Great scientists, artists and philosophers, the ones we all admire and call masters, were and are open minded, curious, and never put barriers between the domains of knowledge, being interested also in music, poetry or art. Poetic or mystical intuitions have influenced scientific intuitions in the past. Vedic texts or the writings of Giordano Bruno mention 'many universes' at the time when there were no big telescopes or complex cosmological models to explain the universe. Nevertheless by means of intuition or a knowledge now lost, the existence of many worlds was predicted. Intuition, a magic word that we often talk about both in science and in art, is what the greatest geniuses have always made treasure of. Dalí, Mondrian, Mozart, Braque, Einstein, Tesla, Poincaré, Feynman… just some of the artists or scientists from a long list who, consciously, have treasured intuition as a precious ally, a most valuable tool for exploring the unknown.

The illustrations presented in this essay are often inherent to the plastic arts on the one hand combined with cosmology and physics on the other, because these are the areas in which I, as an artist with a scientific background, have more experience. Nevertheless, we must not forget that art interacts with all natural sciences: medicine,

biology, chemistry, geology, etc. Everything can be and is a source of inspiration and interaction.

Research, beauty, intuition and curiosity; these are the elements that we find on the shore between art and science and that I will develop in this short essay.

2 Research and Divulgation

More than ten years ago, I left the world of scientific research to dedicate myself to the world of art. Then, not infrequently, it happened that during one of my exhibitions, or conferences, someone would ask: "And now, how do you feel? What's the difference between the life you led before and the life you lead now?" I answered and still answer: "Not much has changed: I was a researcher and now I am still a researcher. It's just the objects and the tools of investigation that, to some extent, have changed." Indeed, the desire to investigate the unknown has remained.

Of course, the scientist focuses on matter, you might say, on the physical world… You might also say 'on what is real'. Right, but I wanted to investigate those parts of reality that cannot be expressed with an equation, but are nonetheless real. The question was: how, as an artist, could I move in this new adventurous and complicated journey of research? What valid tools would be at my disposal? The best compasses, in my opinion, apart from the obvious good use of reason, are beauty and intuition.

In addition, once the results of the research and discoveries have been obtained, they must be communicated, made known and visible. Niels Bohr [3] said: "*We depend on our words… our task is to communicate experience and ideas to others.*" Disclosure is a fundamental and very noble element both for scientists and artists. If a painting remains in a cellar it is of no use to anyone, so it is for a scientific discovery that is kept hidden. The desire to communicate, show and share the results obtained is intrinsic to the world of any scientific or artistic research. It is very noble, it is very beautiful and right; and this act of dissemination can be made very effective and pleasant by the interaction between art and science as presented here.

3 Beauty

Beauty is a very complex subject; entire faculties within universities deal with it. I do not want to turn this essay into a treatise on aesthetics and I would rather leave the concept of beauty to the experience that each one of us has. Indeed, it is something difficult to define in words, elusive and often related to the style of the time. Often, those who attempt to define the concept speak of a certain 'subjectivity', which may ignore paradigms of proportion and perfection, which is frequently presented as a fact. Is beauty an aesthetic experience or an aesthetic emotion as defined by Bell? [4]. Certainly, the human brain plays a critical role in the experience of beauty. Today, also neurobiology is trying to define beauty by identifying the areas of the cortex

that do correlate with experience of visual or musical beauty and pleasure [5], but the author states that the questions raised and unresolved are the same as in philosophy of art and aesthetics [6].

Judith Wechsler, in the introduction to 'Aesthetics in Science', says: "*Scientists talking about their own work and that of other scientists use the terms 'beauty', 'elegance', and 'economy' with the euphoria of praise more characteristically applied to painting, music and poetry*" [7].

This is precisely the point that is close to my heart and that I would like to take to the surface and underline: how beauty is also part of that 'shore' that we find between art and science. And I don't just want to emphasise it where it is most evident, but I also want to find it where we would least expect it. It would be easy to talk about beauty by looking at the works of Raphael, Leonardo or Michelangelo. It would also be easy to talk about beauty, as I said in the introduction, observing an image obtained with the Hubble telescope—a colourful nebula or a shining supernova—to understand that science often works with beauty, but there is much more.

It becomes less obvious, but still acceptable, to speak of beauty regarding the geometric shapes of plants, crystals, or snowflakes. Recognizing in a shell, in a cabbage or in the horns of an antelope the Fibonacci sequence and the Golden Ratio. Man's amazement when faced with the beauty of the forms of nature is still something that does not amaze and that blends together science and art—at least that of the Creator. What is truly astonishing, however, is to find beauty even in what is outside our collective imagination associated with it.

We must not stop at images alone. Beauty is something broader and more deeply rooted, something that can also be found in the notes of a sonata or in the words of a poem. Areas that lack colour and plastic forms, of course, but that can equally arouse in us that emotion produced by the beauty of a painting. But let's go further, even an equation can have these requirements. Niels Bohr, one of the fathers of quantum mechanics, said that "*We must be clear that when it comes to atoms, language can be used only as in poetry. The poet, too, is not nearly so concerned in describing facts with creating images and making mental connections*" [8].

Paul Dirac, another founder of quantum mechanics adds: "*The research worker, in his effort to express the fundamental laws of Nature in mathematical form, should strive mainly for mathematical beauty. He should take simplicity into consideration in a subordinate way to beauty. […] It often happens that the requirements of simplicity and beauty are the same, but where they clash the latter must take precedence.*" And also: "*What makes the theory of relativity so acceptable to physicists in spite of its going against the principle of simplicity is its great mathematical beauty. This is a quality which cannot be defined, any more than beauty in art can be defined, but which people who study mathematics usually have no difficulty in appreciating*" [2].

To this I would also like to add the words of Henri Poincaré, the great French mathematician, who never ceased to amaze us with his wide-ranging visions, open to what transcends the more orthodox version of science: "*It may be surprising to see emotional sensibility invoked à propos of mathematical demonstrations which, it would seem, can interest only the intellect. This would be to forget the feeling of mathematical beauty, of the harmony of numbers and forms, of geometric elegance.*

This is a true aesthetic feeling that all real mathematicians know, and surely it belongs to emotional sensibility. Now, what are the mathematical entities to which we attribute this character of beauty and elegance, and which are capable of developing in us a sort of aesthetic emotion? They are those whose elements are harmoniously disposed so that the mind without effort can embrace their totality while realising the details." [9].

We could go on and on adding names to this already important list, but the message should be clear by now. The wonder of creation has no boundaries and embraces science and art in equal measure. And just as beauty turns out to be a fruit of creativity, it can and must also be used as a compass for the research work that transforms mysterious unknown into knowledge.

4 Intuition

On the shore between art and science there is certainly also the powerful tool of intuition. Intuition, like beauty, is something that could more easily refer to the inspiration of poets, musicians and painters; it could appear as something linked to imagination and fantasy. We are not surprised to read the words of Piet Mondrian: "*Only through intuition does a work rise above more or less subjective expression. Different periods produce different feelings and conceptions, and in each period men differ. Consequently different art expressions even in a single period are not only logical but a tribute to the general development of art. Intuition always finds the way of progress, which is continuous growth toward a clearer establishment of the content of art: the unification of man with the universe.*" [10] Nor those of George Braque: "*I would say that it was 'poetry' which distinguishes the cubist paintings which Picasso and I arrived at intuitively from the lifeless sort of painting which those who followed us tried, with such unfortunate results, to arrive at theoretically.*" [11] Or Mozart saying: "*Thoughts crowd into my mind as easily as you could wish. Whence and how do they come? I do not know and I have nothing to do with it. […] Then my soul is on fire with inspiration […] Then my mind seizes it as a glance of my eye, a beautiful picture or a handsome youth. It does not come to me successively, with its various parts worked out in detail, as they will be later on, but it is in its entirety that my imagination lets me hear it*" [12].

Much more surprising, however, are similar concepts expressed by one of the greatest mathematicians to have walked the planet, Henri Poincaré: "*With pure logic we would have ended up with tautologies; it could not create anything new. Nor could any science be derived from it. In a certain sense, these philosophers are right: to do mathematics, as to do geometry, or any other science, something more than pure logic is needed. To designate this something we have no other word but intuition.*" [9]. The same position, on the side of physics, is held by another giant of the world of science, certainly the best known and most emblematic, Albert Einstein: "*All great scientific achievements arise from intuitive knowledge, that is, from axioms from which deductions are made… Intuition is the necessary condition for the discovery*

of these axioms." [13] *or "Physics is essentially an intuitive and concrete science*" [14].

In Jacques Hadamard's work, 'The Psychology of Invention in The Mathematical Field', there are several examples of other scientists such as Langevin, Helmholtz and Ostwald, who said to have experienced and exploited this tool under the most varied names: illumination, sudden spontaneity, imagination or, indeed, intuition [12].

Intuition, exploited by the great masters of art as well as by the great scientists, together with beauty is also, undoubtedly, a compass to orient oneself and to advance in the infinite sea of research.

A magnificent example of intuition in the world of science, perhaps the first one that really struck me, when I was still a first-year engineering student, is in this anecdote told by Nobel Prize winner Richard Feynman: "*Within a week I was in the cafeteria and some guy, fooling around, throws a plate in the air. As the plate went up in the air I saw it wobble, and I noticed the red medallion of Cornell on the plate going around. It was pretty obvious to me that the medallion went around faster than the wobbling. […] The diagrams and the whole business that I got the Nobel Prize for came from that piddling around with the wobbling plate*" [15].

This is exactly what I mean by 'intuition': an event, a word, an image, a sound… which apparently have nothing to do with the subject of our research, but which provide the key to solving our problem or answering our questions.

5 Synergy Between Art and Science

Albert Einstein [16] says: "*After a high level of technical competence has been reached, science and art tend to merge in aesthetics, plasticity and form.*" This should be the case, but for a long time we have been used to separating science from art as if they were completely distinct fields of human knowledge. It is not easy to establish a starting date for this split, but certainly after the Renaissance, science and art began to distance themselves from each other, which then became extreme and official with the Age of Enlightenment. However, the most enlightened researchers—artists and scientists—have remained in contact without prejudice or presumption. Often mocked by their contemporaries for their open-mindedness, sometimes judged on the verge of naïf, then venerated as icons by subsequent generations. Apart from these not a few bright monads, as I said, the historical lines of science and art separated after the Renaissance and have remained so for a long time. As we have seen, these disciplines have a great deal in common. I believe that the relationship between art and science is today dynamic, synergistic, interactive and constantly evolving. Just as a spinning plate is capable of solving a quantum electrodynamics problem (see Feynman quoted in Sect. 4) so a painting, a piece of music or a poem can become a powerful catalyst for providing answers to old and new human questions. This can happen consciously or unconsciously. However, if we apply this strategy of 'openness to the world' in a conscious way, results are more likely to come. It is also true that

when things are in the air, when the times are ripe, in one way or another, ideas can materialise easily contributing to advancement in various fields of knowledge.

If we take a leap into the second half of the nineteenth century, we encounter the pictorial currents of impressionism followed later by pointillism, cubism, and abstract art when, in the same period, physical theories were born and experimental results were obtained, which led to a vision of matter that was increasingly decomposed. The discovery of the electron and then of the proton confirms that the division of matter does not stop at the atoms, The deconstruction of reality in art begins with the Impressionists, who despite their imprecision of forms still maintain a compact brushstroke, but it became radicalised in the works of Seurat and of the pointillists where each colour acquires its own independence: just like particles. Over the progression of science, claiming the uncertainty principle and more generally the development of quantum physics, artists translated the 'wave-particle' dualism on the canvas into cubist representation of faces and objects seen simultaneously from different perspectives, such as 'Tête de femme II' (1930) by Georges Braque.

What is undoubted is that times are always impregnated with certain energies, certain ideas, and the more the researchers are attentive and open to these, the more they have the possibility of producing new ones to add to the path of human knowledge and evolution.

At the beginning of the twentieth century, the relativity of space and time upset not only Newtonian absolutism, but also our way of seeing reality. Dalí, at that time, painted the 'melting clocks' in the famous work 'The persistence of memory' (1931) being inspired by a melting camembert—the legend tells. But, among so many things that can melt, deform, contract and expand, why the clocks, the symbols of time? Maybe an influence from the world of the new physics of space–time? No doubt, as well, that 'Corpuscular Madonna' (1952) is a reaction to the advances in nuclear physics. And what about 'Galacidalacidesoxyribonucleicacid' (1963), his DNA-inspired work? There is no need to say this to understand that Dalí, like Picasso, was certainly not indifferent to the discoveries of science. Dalí was interested in the physical state of matter and quantum theory, and fascinated by the theory of relativity. He met and dealt with mathematicians (Matila Ghyka and René Thom), physicists (Nobel Prize winners Ilya Prigogine), geneticists (James Watson, the discoverer of the structure of DNA), psychoanalysts (Sigmund Freud, the father of psychoanalysis and Jacques Lacan), film directors and artists (Luis Buñuel, Alfred Hitchcock, Harpo Marx and Walt Disney). His interest in geometry and in three-dimensional vision research led, after the 1960s, to the creation of stereoscopic artworks and from meeting with Nobel Prize Dennis Gabor to the creation of holograms as new artwork expression. As further confirmation, if any were needed, by his bedside (as described by Teresa Vega in her blog [17]) before he died, were these books: 'The Geometry of Art and Life' by Ghyka, 'What is Life?' by Schrödinger, and 'A Brief History of Time' by Stephen Hawking.

Art can certainly represent science, but there are different degrees of representation and different levels of depth. A galaxy, for example, or a supernova, of course, can be painted, but this would be 'naturalism' not different from painters who paint still-life or landscapes. But art can do more. Art can also help science to visualise what is too

complex to express in words and even more complex to be understood by non-experts through equations.

The general theory of relativity is beautiful; its equations are beautiful, but only for experts; and for others? In most cases there is for them neither beauty nor clarity. This marvel of human genius remains remote, excluding them from understanding and expanding their vision of reality. Graphic and artistic representation might come to the rescue. Whether it be Dalí's melted clocks or some more contemporary painting in which space is deformed by the action of stellar and planetary masses, plastic forms help the understanding of the equations in a poetic way.

Surely the opposite has also happened: the fruition of certain works of art like music, poetry, paintings and science fiction stories has undoubtedly influenced scientists. Many scientists have always had an eye for the humanities, arts and music—it is undeniable—just as true artists have always been intrigued by the novelties and discoveries that science and technology has produced.

While reading a book about the Higgs boson, I came across another beautiful example of how science can use art to give substance to certain abstract and complex ideas. In his popular essay, David Blanco Laserna uses an impressionist painting to explain the relativistic quantum field, making an analogy between the small brush strokes of colour and the statistical functions of quantum particles. [18].

There is yet another level in this classification of the interaction between art and science. Before probes and large telescopes were available, imagination played an important role in the representation of the cosmos. The scientific representations done in the paintings of Lucien Rudaux (1874–1947), a French astronomer and pioneer space artist, are a good example of imagination. And it is precisely this kind of imagination that often merges with intuition that the artist can and, in my opinion, must use. Imagination to go beyond, to describe entities of the cosmos that instruments are currently unable to detect or show satisfactorily. Science is by nature more precise, art is softer and in this forest of images and impressions, the right idea or intuition could be hidden.

Dark matter or parallel universes, at the moment, remain only words and predicted entities by theoretical physicists, astrophysicists and cosmologists. If the artist intervenes with his intuition and imagination to realise even very abstract ideas, I believe it is possible that the artistic act can in some way inspire the scientist in the deepest sense of the term, just as happened to Feynman with the spinning plate: unquestionable proof that events external and distant from scientific research can be a catalyst for intuition. Therefore, let's dream that a painting, a symphony or a poem can be a catalyst for the scientific researcher leading, with experiments and theories, to discoveries and answers.

6 Cosmology and Art

Cosmology can be considered a very ancient science, but also an extremely contemporary one. There is talk of cosmogony as early as the Assyrian-Babylonian myths,

at the dawn of human history, but today the mystery and fascination of the cosmos return in a highly scientific guise under the name of cosmology. "*Astronomy is older than physics. In fact, it got physics started by showing the beautiful simplicity of the motion of the stars and planets, the understanding of which was the beginning of physics. But the most remarkable discovery in all of astronomy is that the stars are made of atoms of the same kind as those on the earth*" [19].

The birth of contemporary cosmology can be traced back to the years between the formulation of the theory of general relativity (1915) and 1929, when Hubble published his famous law. It is therefore a very contemporary science precisely because to study the cosmos—the galaxies and the entire universe—conventional optical telescopes were not enough; much more high technology instrumentation, spectrometers, special mirrors, antennas, computers, worldwide team connections and interactions and big data analysis is needed.

Art that interacts with cosmology is certainly contemporary art. As an artist, I try to give form to my intuitions and draw inspiration from what it discovers on a daily basis. The path of my artistic research has been mainly linked to the four elements of the Hellenic tradition: earth, water, air and fire. It started with the earth (Fig. 1), then developed with water (Fig. 2) and fire (Fig. 3) and, from 2017, joining the air to the cosmos with a series of artworks called 'Supernova' (Fig. 4). It has been a long journey from the earth to the stars, but the journey remains a never-ending search on aspects for which I felt affinity and found fascinating both scientifically and artistically: supernova, dark matter, dark energy and multiverse. This work will be displayed in the following sections.

Fig. 1 *THE GOLD OF THE EARTH No. 7.* Acrylic, sand, stones, clay, 23 ¾ kt gold, and canvas on polymeric material, 60 × 60 cm. (© E. Magnani 2012. All rights reserved)

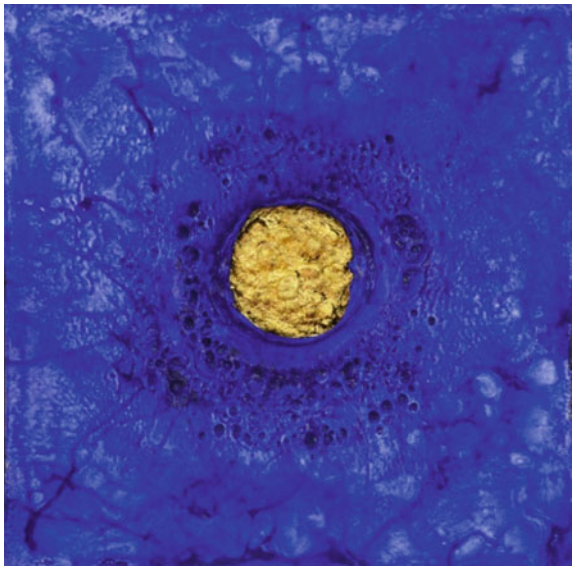

Fig. 2 *NEPTUNE No. 7*. Acrylic, sand, stones, 23 ¾ kt gold on melted polymeric material, 30 × 30 cm. (© E. Magnani 2015. All rights reserved)

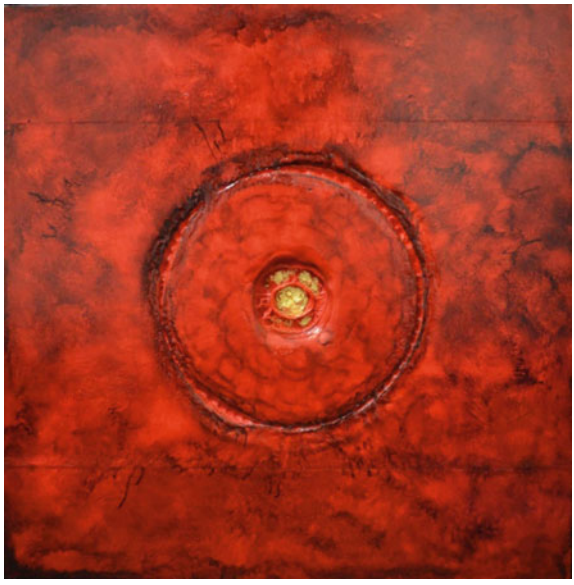

Fig. 3 *RED No. 2*. Acrylic, sand, stones, 23 ¾ kt gold on melted polymeric material, 100 × 100 cm. (© E. Magnani 2015. All rights reserved)

6.1 Supernova

Supernovae are spectacular, very luminous and quite rare stellar explosions that characterise the end of life of massive stars. They produce an emission of radiation that, for short periods, can exceed that of an entire galaxy. The explosion, which

Fig. 4 *SUPERNOVA No. 2*. Acrylic and enamel on paper-aluminium multilayer panel, 100 × 76 cm. (© E. Magnani 2017. All rights reserved)

reaches temperatures of a hundred billion degrees, ejects all the material that makes up the star at a speed of about 30,000 km/s (10% of the speed of light). This creates a bubble of expanding gas that diffuses into the interstellar medium. This released material can lead to the formation of new stars and planets. Stars are true atom factories. From hydrogen, helium and lithium, they are able to create, through nuclear fusion reactions within themselves, all the elements that make life possible. We humans are truly 'stardust' and this is not just a poetic expression. Because of this, in 2017, I began to study the supernova scientifically and to focus my artistic attention on these extraordinary cosmological events. The supernova, a symbol of death, rebirth, and cyclicity of the universe, has an important influence on earth life [20]. That is why supernova has a meaning that goes well beyond cosmology, reminding us that everything is in everything. Supernova unites macrocosm to microcosm, universe to man and, for me, also science to transcendence. A hymn to life with its complexity and its magic.

The artworks of the 'Supernova' series have been made 'at a distance'. In no phase of its creation, the pictorial surface of these works is touched by the artist either with the hands, or with the brushes, or other instruments. These artworks are created using jets of air and water that move the colours on the pictorial surface covered by a veil of water, exploiting the fluid-dynamic effects that, with all the limits of the case, try to simulate the dispersion of elements from the star explosions.

6.2 Dark Matter

My journey to dark matter has been marked by the reading of the book of Cristiano Galbiati 'Le entità oscure—Viaggio ai limiti dell'universo' starting with: "*It all begins in the thirties of the last century, when a Swiss astronomer […] measures for the first time the speed of rotation of a cluster of galaxies with respect to the centre of gravity. […] For a galaxy in the periphery of the cluster, the force of gravity is expected to come from the mass of all the galaxies within the radius of its orbit […] Something is wrong. There is a big discrepancy: the mass is completely insufficient to justify the high speed of the cluster. He realises there is something different, something misunderstood, or hidden in the centre of the cluster. Something unseen. A source of gravitational force that escapes not only the human eye, but also its most refined instruments, a heavy but invisible matter. […] and will never be a source of light, but only and only of gravity. It is dark matter*" [21]. He sums up the essence of one of the hottest and most fascinating topics in contemporary cosmology and physics.

Many artists have dedicated themselves to the representation of dark matter and I too have felt and continue to feel its fascination. An initially unconscious fascination, which, once understood and realised, became more rational.

The works of the 'Supernova Dark Matter' series did not yet have this title when they were created. I merely wanted to integrate a light element in the 'Supernova' series, but for various technical and theoretical reasons, backlighting was no good. Months went by, then, one day, in a shop in the centre of Rome, a craftsman made

me discover a phosphorescent powder that I first used to produce the luminous effect I was looking for. But, there was more to it. Thanks to this, I had the right intuition: phosphorescence is something that cannot be seen in daylight, even if the pigment molecules are present on the paint support. Like dark matter, phosphorescence is there, but cannot be seen. For me, the analogy was enough to work on a dedicated series. The 'Supernova Dark Matter' series was then created having a phosphorescent background made with a photosensitive pigment. In daylight, no one could suspect what is hidden inside the paintings. It is only in the dark that the phosphorescent pigments begin to become visible and emerge amidst the transparent vivid colours of the supernovae. The darker it gets, the more the vivid colours disappear, giving way to a phosphorescent yellow-green light alone, a completely different work from the one visible in daylight. Making visible something that has always been there, but which no one could see in the presence of light (Figs. 5 and 6). The works from the 'Supernova Dark Matter' series are an invitation to change the paradigm of looking at things. It is a simple but effective idea. To see phosphorescence you have to change the paradigm, invert it: remove the light from the environment where the painting is. Furthermore you see something else, different patterns, a unique colour gradient and a painting with different aesthetic appeal.

Here is the message: when, in any field, be it scientific or artistic, we come across a problem that cannot be solved with the traditional means, which we are using or we are accustomed to, it is perhaps time to tackle the issue at its roots. So did Copernicus. In 1500, the astronomical elements known were the same for everyone; but no one could put all the experimental data on the orbits of the planets in an all-encompassing model without making hyperbolic tricks (i.e. the Tycho Brahe's

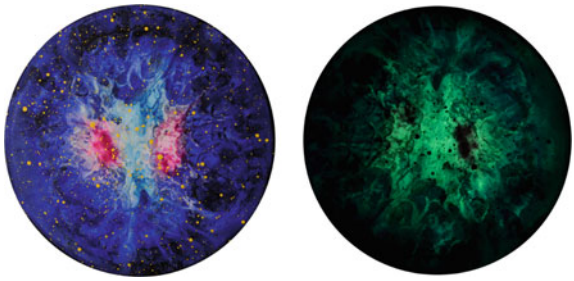

Fig. 5 *SUPERNOVA DARK MATTER R1-19*. Acrylic and phosphorescent pigment on paper-aluminium multilayer panel, Ø 29 cm. *Left*: artwork with light. *Right*: artwork in the dark. (© E. Magnani 2019. All rights reserved)

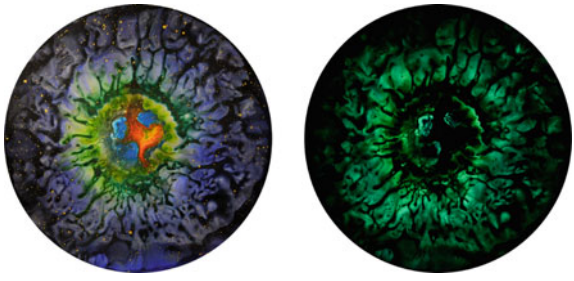

Fig. 6 *SUPERNOVA DARK MATTER R12-19*. Acrylic and phosphorescent pigment on paper-aluminium multilayer panel, Ø 76 cm. *Left*: artwork with light. *Right*: artwork in the dark. (© E. Magnani 2019. All rights reserved)

model). It was enough to change the essential assumption that the Earth was not at the centre of the solar system, but revolved around the sun, to put all the pieces of this puzzle in place with great simplicity. For the sake of completeness, the fascinating thing is that Copernicus was not even the first, if we want to be honest, to place the sun at the centre of the solar system; it was the Greek astronomer and mathematician Aristarchus of Samos, born in 310 B.C., according to Archimede. He considered this possibility supported by Seleucus, according to what Plutarch tells us. But, for about 1400 years, the geocentric vision of Ptolemy reigned.

6.3 Dark Energy

Once again I will use scientific expertise to introduce another mysterious component leading my journey into the unknown: dark energy.

> *"Thanks to the cosmic background radiation, generated at the time of recombination, we are able to precisely determine the composition of the universe to date. Ordinary matter, in the form of protons, neutrons, electrons, contributes only 5% of the energy of the universe, 23% of it is in the form of dark matter and 72% is in the form of dark energy"* [21].

Isn't it intriguing that so much, in the energy balance, is still missing? Isn't it worth focusing on this aspect of reality as well? Trying to play with it also from an artistic point of view? Looking for ways to represent this 'dark entity'? As far as I am concerned, the answer is 'yes'.

For a long time now, there have been certain fairly widespread fluorescent pigments in the art world that are used by artists to achieve the most varied effects for the most varied purposes. Fluorescence, unlike phosphorescence, cannot be seen in the dark: the molecules do not retain energy for a long time, but release it almost immediately. One frequency with which these fluorescent pigments react very well is ultraviolet light. It is easy to find ultraviolet light lamps on the market. The thing that made me turn on a light bulb—pun intended—has been that: ultraviolet light is commonly called 'dark light'. The association was inevitable: dark light - dark energy. And so a new series of works was born exploiting the 'dark light', which is re-emitted when the fluorescent pigment is illuminated.

In the 'Supernova Dark Matter' series, the phosphorescent pigment re-emits the captured light after a while, making itself visible in the dark; in this new series, the fluorescent pigment re-emits the light with a much shorter time and the effect remains visible as long as the external ultraviolet light source is present; the focus is more on light and less on matter (Fig. 7). For painters, light - visible light - has always been a key element, but in terms of intensity and direction; now, as I moved away from the visible spectrum towards the ultraviolet, something new came into play. With the evolution of technology and the uses that can be made of it, both science and art evolve. Science discovers, creates, art follows, applies, and creates again.

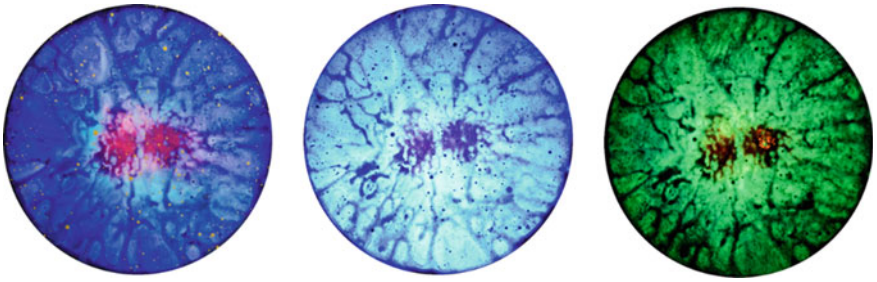

Fig. 7 *SUPERNOVA DARK ENERGY No. 1.* Acrylic, phosphorescent pigment and fluorescent pigment on paper-aluminium multilayer panel, Ø 40 cm. *Left*: artwork with light. *Centre*: artwork with UV light. *Right*: artwork in the dark. (© E. Magnani 2021. All rights reserved)

6.4 Multiverse

In the preface to his book on universes, John Barrow, begins by saying: "*Albert Einstein showed us how to find all possible universes that are consistent with the laws of physics and gravity, and how to reconstruct their past and predict their future. But, actually, finding them was no easy task. Since then, mathematical astronomers and physicists have been striving to solve Einstein's equations and identify those universes*" [22]. These few words are enough to give us an idea of the vastness that opens up before us: truly exciting and challenging prospects.

The concept of the 'multiverse' was first proposed in a rigorous way in 1957 by Hugh Everett III with his many-worlds interpretation of quantum mechanics [23]. The term 'multiverse', however, was coined in 1895 by the American writer and psychologist William James. The idea of parallel universes was also taken up by the American science fiction writer Murray Leinster in 1934 and later by many others, such as Jorge Luis Borges. It was later reaffirmed as a possible consequence of some scientific theories, especially string theory and chaotic inflation or bubble theory. At this point, I cannot fail to point out how intuition on the one hand and/or an ancient knowledge that has been lost on the other hand have already foreseen this possibility. In fact, this many-worlds hypothesis had already been proposed in very ancient Vedic texts (at least 2000 BC), by Democritus (460–370 BC) and other Greek atomists [24]. Can we accept the idea that in a distant past someone knew or had more intuition than we have, although based on different grounds? We are just one of the many explorers trying to understand something in the sea of space and time of our huge and old universe.

6.4.1 Inflationary Multiverse

"According to quantum mechanics, space structure at small scales may fluctuate randomly, giving it somewhat the aspect of a chaotic soapy foam. This leads to the chaotic inflationary

Big Bang theory in which our balloon may get, here and there, formation of kinds of pustules, which, through expansion, get detached thus creating new balloons" [25].

This idea is what prompted me to create a series entirely devoted to the inflationary multiverse. It's an amazing leap forward: we move from stars and galaxies to multiple universes, perhaps, with physics laws completely different from our own. Will there be life, and if yes, in which forms? Perhaps in ways that are unimaginable to us. The existence of multiverses beyond our universe is at present not detectable. Shall we move away from science asking philosophy and art to expand our thinking and our visions? [26].

In my 'Inflationary Multiverse' series of artworks, the colours and shapes of one universe mix on the border with the colours and shapes of the other universes (Figs. 8 and 9). Imagining separate universes that mix in a border area, giving rise to who knows what… These artistic multiverses, with their soft, irregular shapes take up the idea of merging and separating 'bubbles of cosmic foam'—something that instruments are not yet able to 'see'. Will they one day? Today, at least, there is no risk of inaccuracy. Today, art can still afford to use imagination without being disavowed in the search to grasp some small fragments of hidden truths.

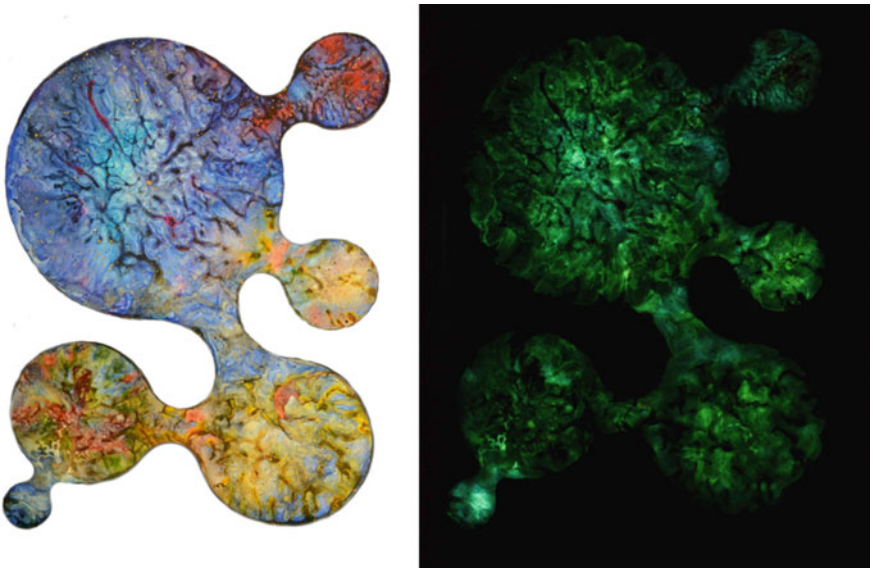

Fig. 8 *INFLATIONARY MULTIVERSE No. 3*. Acrylic and phosphorescent pigment on paper-aluminium multilayer panel, 102 × 76 cm. *Left*: artwork with light. *Right*: artwork in the dark. (© E. Magnani 2019. All rights reserved)

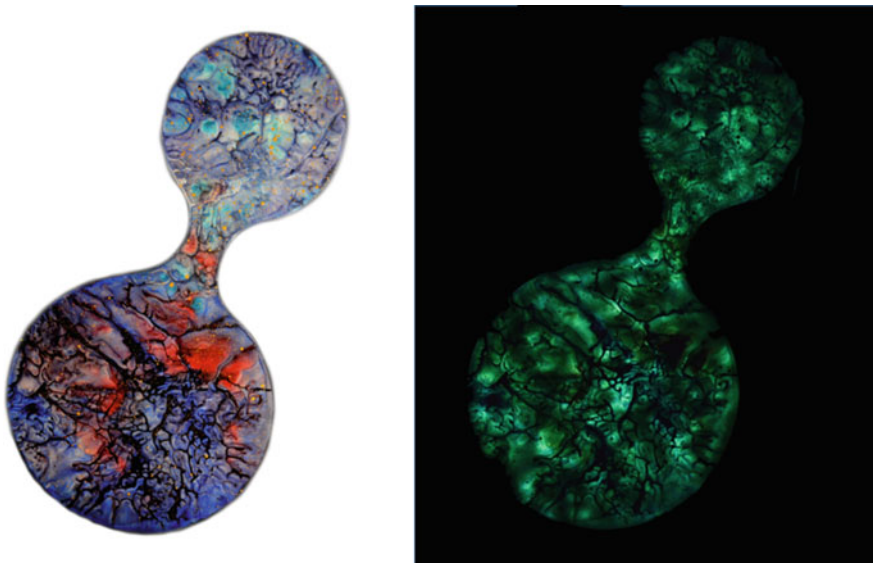

Fig. 9 *INFLATIONARY MULTIVERSE No. 2.* Acrylic and phosphorescent pigment on paper-aluminium multilayer panel, 100 × 56 cm. *Left*: artwork with light. *Right*: artwork in the dark. (© E. Magnani 2019. All rights reserved)

6.4.2 Quantum Multiverse

As anticipated, the concept that there are myriads of worlds was proposed by Hugh Everett III in a rigorous way in his doctoral thesis entitled '*The theory of the universal wave function*', but he never explicitly referred to 'many worlds' [23]. The term was coined by Bryce DeWitt when he began to promote the theory of Everett [27]. For more extensive information you can read 'Many-worlds Interpretation of Quantum Mechanics' [28]. Of course, there is also criticism toward the Many-Worlds Interpretation (MWI) and lot of work has been done on inflationary cosmology with the inflationary model papers of Guth [29, 30] and those of Linde and McCabi, among the many theoretical contribution on the topic [31–34].

In any case, beyond the terminology, Everett's genius lies in considering every superposition element of the universal wave function as 'a world', with all the astonishing things that can follow. To sweeten this concept, which arises from quantum theory, let us think of a cake with layers, parallel and having a circular base. The cake represents the wave function of the generic state of 'everything that exists'. Each layer of the cake corresponds to a superposed element of the wave function, which we can also call 'world' or 'universe'. The cake is so varied that each layer is different from the others. These universe-layers are infinite in number and each contains at least one observer who observes different events, but always only one event at a time. Furthermore, in the multiverse, every time a choice is made, the others are also realised, because our doubles in the parallel universes make all of

them; and this applies to all observers and to any event! Infinite power infinity! Here is a taste of parallel universes.

A truly stimulating concept that would be difficult not to attempt to represent artistically. Thus, from this, my series of works 'Quantum Multiverse' was born. The metaphor of the cake layers is also what makes the connection with the artistic representation I decided to make (Fig. 10). The variegated circles that appear in my works, randomly arranged on a black background, can be seen as some of the layers of this cake, taken and thrown into the cosmic void. Of course, there are many limits: infinity is a bit difficult to represent. Inevitably, there is a reference and also an homage to Wassily Kandinsky's 1926 work 'Several Circles'. It is curious that one of the possible interpretations of this work—as the title is very generic—is precisely cosmological, or at least astronomical. Some have seen in it planets and stars arranged in more or less dense clusters, including an eclipse. In my multiverse paintings, however, a further element appears: tiny golden particles that are the only ones allowed to circulate in this cosmic void between universes. After all, if the universes are to be coordinated, it is also possible that there are messengers that connect everything. Could they be the gravitons? Or maybe the tachyons, those particles with imaginary masses that speed faster than light? Is imagination going very far, maybe too far?

7 Interaction Between Art and Science—Inside and Outside the Work of Art

When we talk about the interaction between art and science, we have to distinguish how and where the interaction takes place. The interaction can take place both inside and outside the work. Let me explain.

When the artist creates the work, if the theme is scientific, we have an interaction between art and science within the work itself. Then, how to show this interaction? The science-themed artwork is a perfect element of union between art and science. If we want to bring these two worlds closer together, we need to start by exposing and emphasizing the common ground, showing to the general public as well as professionals that these two worlds really do have a lot in common. Once we understand and accept this, extension will happen naturally. Exhibiting a Rubens at CERN and the equations of General Relativity at MoMA (Museum of Modern Art), beyond the provocative gesture, would not make the link between art and science explicit. These two worlds would continue to appear to the collective imagination separate and distant. Science-themed artworks, instead, can be reasonably exhibited in both these two locations. This is why works of art that talk about science are so important today. They are the passe-partout to open the vision and trigger the dialogue between art and science.

This interaction between art and science in public places is what we have achieved between 2019 and 2020. The artworks on supernova, multiverse, and dark matter

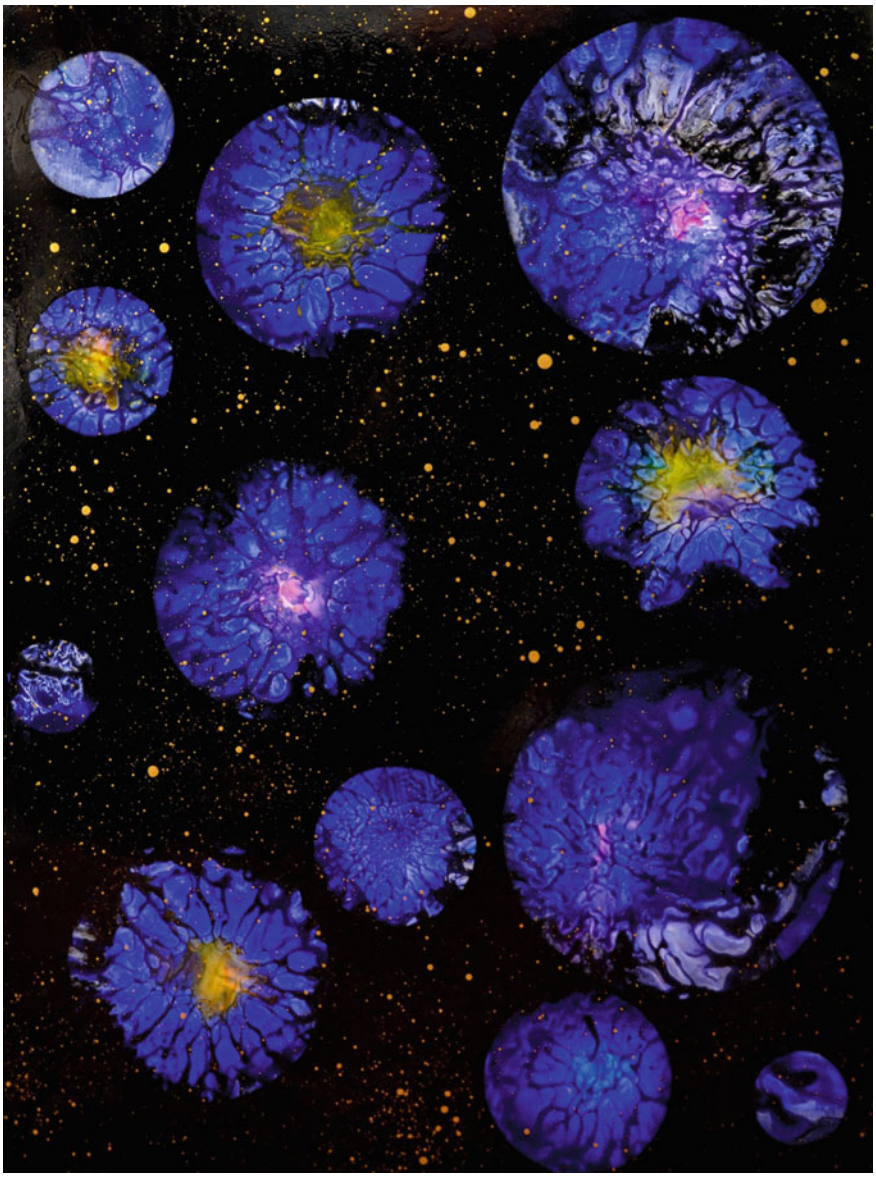

Fig. 10 *QUANTUM MULTIVERSE No. 2.* Technique: acrylic and enamel on paper-aluminium multilayer panel, 100 × 76 cm. (© E. Magnani 2019. All rights reserved)

have been presented at the European Organisation for Nuclear Research (CERN), in Geneva, at the Gran Sasso Science Institute (GSSI), and at the Gran Sasso National Laboratory (LNGS) in L'Aquila.

7.1 Searching the Unknown

The travelling exhibition entitled '*Enrico Magnani - Searching the unknown'* curated by Marilena Streit-Bianchi was shown in the framework of CERN Staff Association Exhibition in July 2019 at CERN, in Geneva, Switzerland (Figs. 11 and 12) and during the CERN Open Day at the CMS (Compact Muon Solenoid) experiment premise and subsequently in September of the same year at GSSI, on the occasion of the 105th National Congress of the Italian Physical Society (SIF) and of the European Researchers' Night in L'Aquila, Italy (Fig. 13). We could on many occasions highlight to young students and to the public the points of union between art and science, two fields that in the collective imagination often appear clearly separated. The 'Supernova Dark Matter' series was a way to attract the attention of a public unfamiliar with the scientific world and to talk in a visual way about one of the most fascinating themes in contemporary physics and cosmology as the exhibition was complemented by attractive posters and a catalogue with a section written by scientists in a language accessible to the wider public 'Art Meets Science—Perspectives on Dark Matter' [35] and the section containing the artworks 'Searching the Unknown—The Dark Matter Collection'. In addition to have been able to create a good synergistic and stimulating effect, I cannot forget the importance of the inbreeding of knowledge and the dissemination of information on the status of art in an important field of research where several thousand researchers and engineers worldwide are working to advance our understanding of the universe we are living in.

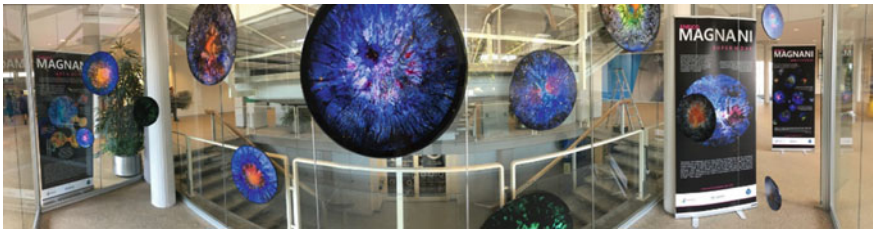

Fig. 11 Installation for the exhibition 'Enrico Magnani—Searching the Unknown' CERN, Geneva, Switzerland, 16–26 July, 2019. (Photo © E. Magnani 2019. All rights reserved)

Fig. 12 The artist at work during the installation of the exhibition 'Enrico Magnani—Searching the Unknown' CERN, Geneva, Switzerland, July 2019. (Photo © K. Petrick—E. Magnani 2019. All rights reserved)

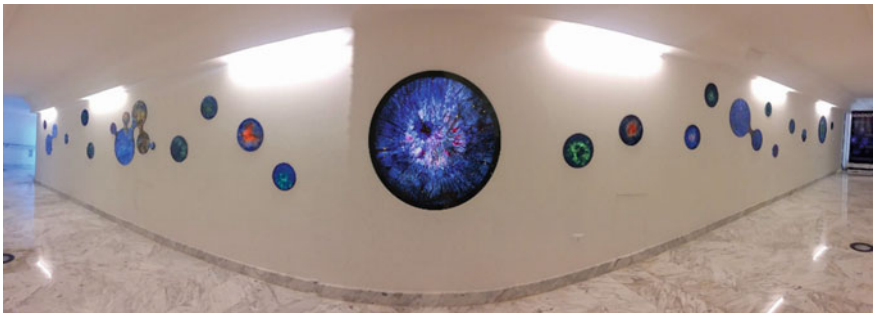

Fig. 13 The exhibition 'Enrico Magnani—Searching the Unknown' at GSSI (Gran Sasso Science Institute), L'Aquila, Italy, 23–27 September 2019. (Photo © E. Magnani 2019. All rights reserved)

7.2 Quintessence

In 2019, the LNGS commissioned an artistic representation of astro-particle physics to remain permanently on their premises: a visual emblem of the themes and research that had been and is being carried out there, particularly on dark matter (with the experiments: Dama, Cosinus, Cresst, Darkside, NEWSdm, Sabre, Xenon) and neutrinos (with the experiments: Borexino, Cuore, Cupid-0, Gerda, LVD). The work was created and permanently installed in 2020. This gave me the opportunity to bring the acquired scientific knowledge into aesthetic and plastic representation incorporating reference to myths, to the creation of matter, and the magic and alchemical process of art creation.

'Quintessence' consists of five parts that are placed on five different walls of the LNGS Research Division. They present portions of the universe with stars, supernovae and planets. Four walls contain rectangular installations of 5×1.35 m (Fig. 15) and to each is associated an element of the Hellenic tradition and one of the four fundamental forces (gravitational force—earth, strong nuclear force—water, weak nuclear force—air, electromagnetic force—fire). In the inner core of the supernovae represented in each of the four walls, as shown in the following figures, the elements are represented with the archetypal colours of the tradition: yellow and brown for earth (Fig. 15a), turquoise and green for water (Fig. 15b), blue and violet for air (Fig. 15c), orange and red for fire (Fig. 15d). Black is associated to the cosmic void. The fifth wall houses a circular installation with a diameter of three metres (Fig. 16) dedicated to the mysterious 'dark matter', metaphorically represented by phosphorescent works that glow in the dark completely changing their appearance and making visible the invisible.

But why I called the complete artworks installation 'Quintessence' and the background underlying the creation of that work need to be explained more in detail.

In Aristotle's physics, the heavenly bodies, above the sphere of the moon, consist of a Quintessence of eternal and incorruptible nature. It is undoubtedly something high, which transcends what goes beyond reality and earthly vision. It is that entity

Fig. 14 The artist at work during the installation of 'Quintessence' at LNGS (Gran Sasso National Laboratory), L'Aquila, Italy. October 2020. (Photo © K. Petrick—E. Magnani 2020. All rights reserved)

which synthesises the four elements, the perfect synthesis, the original idea, the one, the god, the source. The four elements of the Greek tradition cannot fail to remind us of the four states of matter—the most known, but there are more—of which the universe is composed, and also of the four fundamental forces of which man has long been seeking the unification. The first force from which all others originate is the Quintessence. We are not far off, three forces have already been unified, but we are still missing one to be able to prove the existence in the very early universe of a grand unification epoch in which the fundamental interactions were united. The name of my installation is an omen to cosmology and fundamental research.

7.2.1 Supernova—Life, Death and Rebirth

Four of the five walls of the 'Quintessence' installation contain artworks belonging to the 'Supernova' series. At LNGS the experiments are underground. Why, then, represent supernovae that are in the sky? Because the main studies at LNGS are on neutrinos and supernovae are huge sources of these particles; when a star explodes, almost 99% of the energy emitted is in the form of neutrinos. These elusive particles therefore have the possibility of arriving on earth and even underground. In particular, the Large Volume Detector (LVD) is dedicated to the search for gravitational collapse neutrinos that come from supernovae; this is the scientific side. On the philosophical side, as mentioned in the dedicated section, the supernova is at the origin of life.

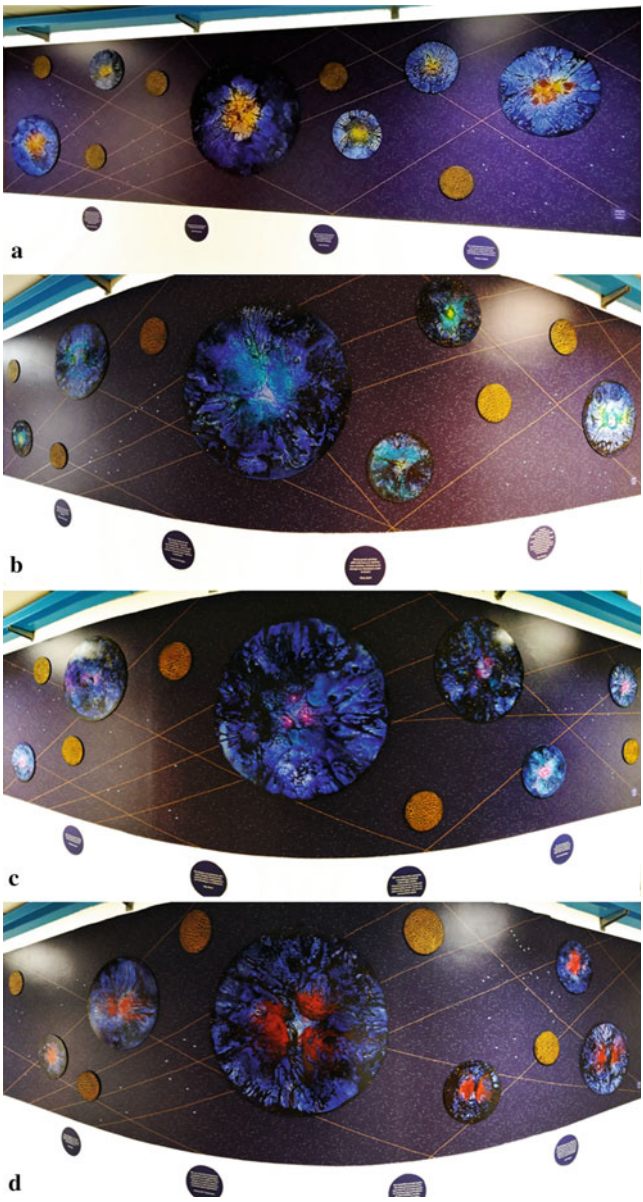

Fig. 15 **a** *QUINTESSENCE—EARTH,* **b** *QUINTESSENCE—WATER,* **c** *QUINTESSENCE—AIR,* **d** *QUINTESSENCE—FIRE.* Each wall is composed by six artworks of the 'Supernova' series and five artworks with cretto, on a sky composition designd by the artist and printed on PVC, 500 × 135 cm. Photos b, c and d taken with fisheye lens. (© E. Magnani 2020. All rights reserved)

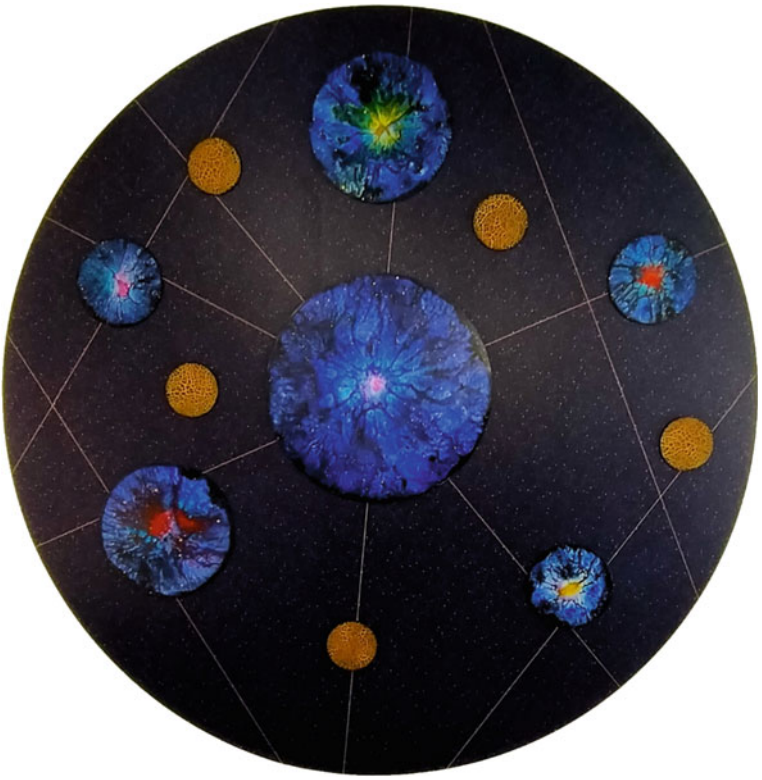

Fig. 16 *QUINTESSENCE.* Six artworks of the 'Supernova Dark Matter' series and five artworks with cretto, on a sky composition designed by the artist and printed on PVC, Ø 300 cm. (© E. Magnani 2020. All rights reserved)

Thus, the death of a star, in cosmic times, gives birth to life: a great symbol of the cyclical nature of the universe.

7.2.2 The Constellations—Astronomy and Astrology, Science and Myth

The constellations can be seen in completely different ways by astronomy and astrology. They are not only the celestial bodies studied by physics and cosmology, but also mythological entities that leave room for imagination, predicting and influencing the destiny of man. Until the seventeenth century no distinction existed between astrology and astronomy. This is the reason that prompted me to integrate them into Quintessence. Each composition that appears on the first four walls of the installation is associated with an element: Earth, Water, Air and Fire. The twelve constellations of the zodiac are also traditionally associated with these four elements. On the first wall, dedicated to Earth, Taurus, Virgo and Capricorn appear;

in the second, dedicated to Water, Cancer, Scorpio and Pisces appear; in the third, dedicated to Air, Gemini, Libra and Aquarius appear; in the fourth, dedicated to Fire, Aries, Leo and Sagittarius appear.

7.2.3 The Earth and the Planets—Challenging Geo-Neutrinos

The cosmos is not only populated with stars, but also with planets. Furthermore, the Earth is also a source of geo-neutrinos, antineutrinos created in the nuclear decay of some of the elements present in the earth interior and experiments studying them are located under the Gran Sasso mountain. The works representing the planets were made with clays which, when drying, spontaneously create a crack.

7.2.4 Neutrinos—The Elusive Travellers of the Universe

When I started thinking about setting up the installation at LNGS, one of the first challenges I had to face was how to represent neutrinos naturally produced in the sun and by supernovae explosions whose physic is extensively studied there. Neutrinos pass through everything they encounter: space, stars, planets, earth. They are so light, elusive and ethereal that they led Pauli to say: "*I have done a terrible thing, I postulated the existence of a particle that cannot be detected* [36]." Gold is a precious colour, a symbol of wealth, elevation and transcendence. Thin golden straight lines that originate from celestial bodies and let themselves be captured only in exceptional cases are the artistic representation that I have given to such precious particles in the LNGS installation.

7.2.5 Dark Matter—Phosphorescence as a Change in Paradigm

The works on the fifth and last wall belong to the 'Supernova Dark Matter' series and were created with the same technique as the 'Supernova' series, adding a background of phosphorescent pigment whose pictorial effects, invisible with light, are revealed in the darkness. The aesthetics of the works in the dark changes unexpectedly with the appearance of new colours and new shapes, as already said a metaphor for dark matter, one of the main protagonists in the world of cosmological research that no one has yet managed to capture. This pictorial emblem of the complexity and relativity of the universe is an invitation to refine the tools of investigation and to always push knowledge beyond by overcoming the investigative barriers and prejudices that bind the thought.

7.2.6 The Fibonacci sequence—The Numbers of the Beauty of Nature

The Fibonacci sequence (0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89…) is probably the most famous number sequence. Why? Because if we observe nature, if we 'measure' nature, we realise that it often hides this numerical sequence or something that originates from it in a very elegant way. Mathematics, biology, art and beauty find a perfect synthesis in the Fibonacci numbers. It also contains the equally famous Golden Ratio ($\phi = 1{,}618\ldots$) on the basis of which wonderful buildings and masterpieces were built. In the creation of Quintessence, the Fibonacci sequence is the basis of the size and the number of works that each wall contains. For reasons of composition and space, the choice was as follows: one work with a diameter of 89 cm, two works with a diameter of 55 cm, three works with a diameter of 34 cm and five works with a diameter of 21 cm. Furthermore, the golden lines that cross the space, an artistic representation of neutrinos, originate from points that are all proportional to the Golden Ratio.

7.2.7 The Quotes—Words Are Part of the Work.

The five walls on which the installation 'Quintessence' develops integrate a series of quotes by great scientists who have contributed to the evolution of humanity through their ideas and their research. I have selected these particular quotes not to forget how much the giants of science had a 360-degrees gaze on reality, without ever excluding mystery, intuition, art, beauty, freedom of thought, the feeling of a higher unity and the courage to break the old paradigms when faced with the evidence.

8 Conclusions

With this quick flight along the shore between art and science, I wanted to highlight many of the aspects common to these two disciplines. The topic is certainly very broad and very stimulating and I limited it to the sector I have been attracted to in the past years: cosmology. Each section treated would deserve to be developed further in order to identify all the repercussions that this inter-disciplinary interaction could have. It would be enough if this short essay is welcomed as a spark to ignite the desire to explore these issues and expand our common visions, a spark to activate that openness to the inner and outer worlds that allows us to study and experience life in its entirety, a spark to ignite that interest towards the mysterious drive that characterises every research discipline, being it scientific or artistic. Curiosity, a powerful engine that drives humanity on its quest of the unknown.

References

1. A. Einstein, *My Credo* (1932). https://www.einstein-website.de/z_biography/credo.html
2. P.A.M. Dirac, The relation between mathematics and physics. *Proceedings of the Royal Society* (Edinburgh), vol. 59, pp. 1938–1939, Part II 122–129 (1939) and https://www.azquotes.com/quote/574156 and https://www.azquotes.com/quote/712210
3. N. Bohr, *Philosophy of Science*, vol. 37. p. 157 (1934) and also R.G. Newton, *The Truth of Science: Physical Theories and Reality*, p. 176 (1997)
4. C. Bell, *Art*. London: published by Chatto & Windus. Art from Project Gutenberg online-books.library.upenn.edu (1914)
5. S. Zeki, Notes towards a (Neurobiological) definition of beauty. *Gestalt Theory*, vol. 41, N. 2, pp. 107–112 (2019). ISSN 2519; DOI: https://doi.org/10.2478/gth-2019-0012 V
6. G. Graham, *Philosophy of the Arts* (Routledge, An Introduction to Aesthetics, 2005)
7. J. Wechsler, *On Aesthetics in Science* (The MIT Press, Cambridge, MA, 1978), p. 1. ISBN: 0–262–23088–7. DOI: https://doi.org/10.177/016224397800300379
8. N. Bohr, Discussions about Language. *Quoted in Defense Implications of International Indeterminacy* (1972), by J. Robert, P. Pranger, p. 11, and Theorizing Modernism : Essays in Critical Theory (1993), by Steve Giles, p. 28 (1933)
9. H. Poincaré, The foundation of Science: Science and hypothesis, The value of Science, Science and Method. Translated by Halsted G.B. released on May 17, 2012. Gutenberg.org/files/39713/39713-h/39713-h.htm. Science and method. Originally published in 1904. The value of science. Originally published in 1905 (1908).
10. P. Mondrian, A New Realism, 1940's. A New Realism, 1943–1945. p. 17 (1945). https://quotepark.com/quotes/1738737-piet-mondrian-only-through-intuition-does-a-work-rise-above-more
11. J. Richardson, G. Braque, 1946–1963, interview with John Richardson (1957). https://quotepark.com/quotes/1807467-georges-braque-i-would-say-that-it-was poetry-which-distinguish/
12. J. Hadamard, *The Psychology of Invention in the Mathematical Field*, (Princeton University Press, 1945). or Ed. Isha Books 1 Janvier 2013. ISBN-10:9333158928 and ISBN-13:978–9333158923
13. A. Moszkowski, H. L. Brose, H. LeRoy Finch. Conversations with Ein-stein. New York Horizon Press (1970)
14. A. Einstein, Lettre à Maurice Solovine. Gauthier-Villars Editeurs, Imprimeurs, Libraires (1956). Paris 1956 and K. Brecher, (1979) Albert Einstein: 14 March, 1879—18 April, 1955 A guide for the perplexed. Nature **278**, 215–218 (1979). https://doi.org/10.1038/278215a0
15. R.P. Feynman, Surely you're joking Mr. Feynman!, Adventures of a Curious Character Ed. (Norton W.W. & Company, 1985). ISBN: 0–393–01921–7
16. A. Einstein, Einstein Archive 33–257 in A. Henderson (1923). Durham Morning Herald (21 Aug 1955) and A. Calaprice, (1996) The Quotable Einstein p. 171.
17. Vega T., Post (2022). https://madridcoolandcultural.com/salvador-dali-art-science/MadridCool&Cultural
18. D. Blanco Laserna, Il bosone di Higgs- La particella che dà sostanza a tutte le cose. Ed. RBA ISSN: 2421–3993 translated by Verardi E. Della Giovampaola G. and Rambaldi R. Italian version (2016)
19. R.P. Feynman, Feynman lectures on physics. (Feynman, Leighton, Sands.) *The Relation of Physics to Other Sciences*, vol. 1, 3–4 (1961). Astronomy http://feynamnlectures.caltech.edu
20. H. Svensmark, Supernova rates and burial of organic matter. Geophys. Res. Lett. (2022). https://doi.org/10.1029/2021GL096376
21. C. Galbiati, (2018) Le entità oscure—Viaggio ai limiti dell'universo. Feltrinelli editore 4th June 2018. ISBN-10: 8807492393 ISBN-13: 978–8807492396
22. J.D. Barrow, *The Book of Universe* (Bodley Head Publisher, 2011). ISBN-10: 1847920985 and ISBN-13: 978–1847920980
23. H. Everett III., Relative state formulation of quantum mechanics. Rev. Mod. Phys. **29**, 454–462 (1957)

24. E. Danezis, E. Theodossiou et al., The cosmology of democritus. Bulgarian Astron. J. **13**, 140–152 (2010)
25. J. Heidmann, The plurality of inhabited worlds. From Bruno to the inflationary big bang, in *Immagini del cosmo*, ed. by A. Omizzolo A. Cosmologia, filosofia, arte. Ed. Il Poligrafo, 1st Jan 2004. 147 ISBN-10: 8871151682 and ISBN-13: 978–8871151687
26. G. Cleaver, *Multiverse theories: philosophical and religious perspectives* (Oxfordre.com Published on line, Oxford Research Encyclopaedia of Religion, 2019). https://doi.org/10.1093/acrefore/9780199340378.013.157
27. B.S.M. DeWitt, Quantum mechanics and reality. Phys. Today **23**(9), 30–35 (1970). https://doi.org/10.1063/1.3022331
28. Stanford Encyclopedia of Philosophy, Many-Worlds Interpretation of Quantum Mechanics. First published on 24th March 2002, revision 5th Aug.ust 2021. Plato.stanford.edu (2021)
29. A.H. Guth, The inflationary Universe: The Quest for a New Theory of Cosmic Origin. *Basic Books* (1997). ISBN: 978–0201328400
30. A.H. Guth, Eternal Inflation and its implications. J. Phys. A **0**(25), 6811–6826 (2007). arXiv: hep-th/0702178. https://doi.org/10.1088/1751-8113/40/5/S25
31. A.D. Linde, Particle physics and Inflationary Cosmology. Harwood Publisher, Chur, Switzerland Berlin Heidelberg **2008**, 1–54 (1990)
32. A.D. Linde, Inflationary cosmology. 22nd IAP colloquium 'Inflation+25' Paris June 2006 Lect. Notes Phys. **738**, 1–54 (2007). arXiv: 0705.0164v2 DOI https://doi.org/10.1007/978-3-540-74353-8_1
33. A.D. Linde, *Inflation and Quantum Cosmology* (Publisher Academic Press, 14 Nov. 2012). ISBN-10: 0124336930 and ISBN-13: 978–0124336933 18 January 2006
34. G. McCabe, *Possible Physical Universes* (2008). arXiv:gr-qc/0601073v1
35. E. Magnani, L. Cifarelli, V. Dehò, L. Verde, M. Streit-Bianchi, A. Balbi, L. Alvarez-Gaumé, T. Camporesi, M. Massimi, B. Bressan, Searching the Unknown The Dark Matter Collection. When Art meets Science Perspectives on Dark Matter. Ed. Kaiti expansion (2019)
36 W. Pauli, statement of 1930, quoted by Frederick Reines, in his Foreword to Spaceship Neutrino (1992) by Christine Sutton, p. xi, Cambridge University Press, Paperback, ISBN: 9780521367035