



LUND UNIVERSITY

Development of New Approaches to ATLAS Detector Simulation and Dark Matter Searches with Trigger Level Analysis

Marcon, Caterina

2021

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Marcon, C. (2021). *Development of New Approaches to ATLAS Detector Simulation and Dark Matter Searches with Trigger Level Analysis*. Lund University.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



Development of New Approaches to ATLAS Detector Simulation and Dark Matter Searches with Trigger Level Analysis

CATERINA M. L. MARCON

FACULTY OF SCIENCE, DEPARTMENT OF PHYSICS | LUND UNIVERSITY



Development of New Approaches to ATLAS Detector Simulation and Dark Matter Searches with Trigger Level Analysis

Development of New Approaches to ATLAS Detector Simulation and Dark Matter Searches with Trigger Level Analysis

by Caterina M. L. Marcon



LUND
UNIVERSITY

Thesis for the degree of Doctor of Philosophy
Thesis advisors: Dr. Oxana Smirnova
Faculty opponent: Dr. Claudio Grandi

To be presented, with the permission of the Faculty of Science of Lund University, for public criticism in the
Sal B lecture hall at the Department of Physics on Friday, the 29th of October 2021 at 13:15.

Organization LUND UNIVERSITY Department of Physics Box 118 SE-221 00 LUND Sweden		Document name DOCTORAL DISSERTATION	
		Date of disputation 2021-10-29	
Author(s) Caterina M. L. Marcon		Sponsoring organization	
Title and subtitle Development of New Approaches to ATLAS Detector Simulation and Dark Matter Searches with Trigger Level Analysis			
Abstract <p>Elementary particles and their interactions are successfully described by the Standard Model of particle physics (SM). However, it has been observed that extensions Beyond the Standard Model (BSM) are required to account for a large part of yet undiscovered particles and interactions, such as Dark Matter (DM).</p> <p>To advance the knowledge of the SM and to pursue DM discoveries, CERN has the ambitious plan of further increasing the Large Hadron Collider's (LHC) energy and luminosity, thus reaching unprecedented event rates in the field of collider physics.</p> <p>This thesis is divided in three parts, dealing with some of the most challenging aspects of the ATLAS experiment at the LHC present and future activities. After a thorough review of the SM, BSM physics is outlined, with particular attention to DM searches. The second part of this work addresses the issue of coping with the foreseen event rates of the High-Luminosity LHC (HL-LHC) phase. Indeed, optimizations of the existing Geant4 simulation codes are a crucial step to alleviate the need for new and expensive hardware resources. With the objective of improving the efficiency of the simulation tools, an extensive study on different compilers, different optimization levels and different build types is presented. In addition, a preliminary investigation on the geometry description of the ATLAS Transition Radiation Tracker (TRT) modules is discussed. The last part of the thesis covers the DM searches carried out by the ATLAS Trigger-object Level Analysis (TLA) group. These searches are based on the analysis of the invariant mass spectrum of di-jet events and, during LHC Run 2, have been performed at energies in the 450 – 1800 GeV range (integrated luminosity up to 29.3 fb⁻¹ and center of mass energy of 13 TeV). After a review of the TLA studies, a preliminary investigation on the performance of Bayesian and Frequentist statistical tools is presented. In particular, the attention is focused on the interpretation and handling of systematic uncertainties both on background and DM signals. This is of particular importance in the process of finding localized excesses, which can indicate the existence of DM signals, and setting limits on the DM event cross sections.</p>			
Key words particle physics, computing, Geant4, optimizations, dark matter, di-jet resonances, ATLAS, LHC			
Classification system and/or index terms (if any)			
Supplementary bibliographical information		Language English	
ISSN and key title		ISBN 978-91-7895-883-2 (print) 978-91-7895-884-9 (pdf)	
Recipient's notes		Number of pages 202	Price
		Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources the permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature 

Date 2021-06-15

Development of New Approaches to ATLAS Detector Simulation and Dark Matter Searches with Trigger Level Analysis

by Caterina M. L. Marcon



LUND
UNIVERSITY

Cover illustration front: G. Balla, *Forme grido Viva l'Italia*, 1915.

© Caterina M. L. Marcon 2021

Faculty of Science, Department of Physics

ISBN: 978-91-7895-883-2 (print)

ISBN: 978-91-7895-884-9 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2021



Abstract

Elementary particles and their interactions are successfully described by the Standard Model of particle physics (SM). However, it has been observed that extensions Beyond the Standard Model (BSM) are required to account for a large part of yet undiscovered particles and interactions, such as Dark Matter (DM).

To advance the knowledge of the SM and to pursue DM discoveries, CERN has the ambitious plan of further increasing the Large Hadron Collider's (LHC) energy and luminosity, thus reaching unprecedented event rates in the field of collider physics.

This thesis is divided in three parts, dealing with some of the most challenging aspects of the ATLAS experiment at the LHC present and future activities. After a thorough review of the SM, BSM physics is outlined, with particular attention to DM searches. The second part of this work addresses the issue of coping with the foreseen event rates of the High-Luminosity LHC (HL-LHC) phase. Indeed, optimizations of the existing Geant4 simulation codes are a crucial step to alleviate the need for new and expensive hardware resources. With the objective of improving the efficiency of the simulation tools, an extensive study on different compilers, different optimization levels and different build types is presented. In addition, a preliminary investigation on the geometry description of the ATLAS Transition Radiation Tracker (TRT) modules is discussed. The last part of the thesis covers the DM searches carried out by the ATLAS Trigger-object Level Analysis (TLA) group. These searches are based on the analysis of the invariant mass spectrum of di-jet events and, during LHC Run 2, have been performed at energies in the 450 – 1800 GeV range (integrated luminosity up to 29.3 fb^{-1} and center of mass energy of 13 TeV). After a review of the TLA studies, a preliminary investigation on the performance of Bayesian and Frequentist statistical tools is presented. In particular, the attention is focused on the interpretation and handling of systematic uncertainties both on background and DM signals. This is of particular importance in the process of finding localized excesses, which can indicate the existence of DM signals, and setting limits on the DM event cross sections.

A mio nonno.

Contents

Abstract	vii
List of publications	iv
Acknowledgements	v
Popular summary in English	vi
I Physics at the Large Hadron Collider	I
1 The Standard Model	5
1.1 Introduction to fundamental concepts	5
1.1.1 Standard Model Lagrangian	6
1.1.2 Symmetries	7
1.1.3 Fermions	9
1.1.4 The SM forces	13
1.2 Phenomenology of QCD	18
1.2.1 Evolution of a hadronic collision	18
1.2.2 Jets	22
1.2.3 The PDFs	23
1.2.4 Factorization theorem for hadronic cross sections	24
2 Searching for Dark Matter	27
2.1 Evidences for Dark Matter: more than a century of research	28
2.1.1 The early stages of Dark Matter studies	28
2.1.2 Rotation curves	29
2.1.3 Gravitational lensing and bullet cluster	29
2.1.4 Simulations	32
2.1.5 Cosmological Microwave Background	33
2.2 Dark Matter candidates	34
2.2.1 Neutrinos	35
2.2.2 Axions	36
2.2.3 Weakly Interacting Massive Particles	36
2.2.4 MACHOS	38
2.2.5 MOND	39
2.3 WIMPs searches	39

2.3.1	Direct search	40
2.3.2	Indirect search	42
2.3.3	Colliders search	43
3	LHC and the ATLAS experiment	47
3.1	The Large Hadron Collider	47
3.1.1	Acceleration mechanism of protons	48
3.1.2	Luminosity and pile-up	49
3.1.3	Run schedule	51
3.2	The ATLAS detector	52
3.2.1	The philosophy of the ATLAS design	52
3.2.2	Requirements	54
3.2.3	The ATLAS detector sub-systems	55
3.2.4	Trigger system	66
3.2.5	Data Handling	67
3.2.6	Monte Carlo Simulation	68
II	Optimizing the ATLAS simulation workflow	71
4	Monte Carlo simulations and software optimizations	75
4.1	Monte Carlo simulations with Geant4	75
4.1.1	The Monte Carlo method	75
4.1.2	Geant4 structure	78
4.1.3	Geometry and material description	78
4.1.4	Physics lists and modules	79
4.1.5	Geant4 simulation workflow	80
4.1.6	Data extraction	81
4.2	Fundamentals of software optimization	82
4.2.1	Abstraction in computer systems	82
4.2.2	Compilers and linkers	86
5	Performance optimization of simulation workflows	93
5.1	Methodology	94
5.1.1	Details of the Geant4 benchmark simulation	94
5.1.2	Runtime geometry representations	96
5.1.3	Compiler's optimization levels	99
5.2	Impact on the simulation execution time	101
5.2.1	Studies with different build types	101
5.2.2	Investigations on the effects of different optimization flags	106
5.2.3	Different primary particles	107
5.3	Conclusions and outlook	III
5.3.1	Build types	III

5.3.2	Optimization flags	III
5.3.3	Evolution of GCC compilers	II2
5.3.4	Primary particles	II3
5.3.5	Final remarks	II3
6	Optimization of detector geometry description	II5
6.1	Sources of geometry descriptions	II5
6.2	Assessment of GeoModel accuracy	II6
6.3	Implementation of new shapes	II20
6.4	Conclusions and outlook	II23
III	Investigations of statistical tools for Trigger Level Analysis	II25
7	Dark Matter searches at colliders using di-jets	II29
7.1	Dark Matter signatures at LHC	II30
7.1.1	Theoretical framework	II30
7.1.2	Experimental signatures	II32
7.2	Jets reconstruction	II34
7.2.1	Input to jet algorithms	II35
7.2.2	Jet finding algorithms	II37
7.2.3	Calibration	II39
7.3	Di-jet analyses workflow	II41
7.3.1	Probability definition: frequentist vs Bayesian	II42
7.3.2	Search Phase	II43
7.3.3	The background estimate	II44
7.3.4	Validating the background-only hypothesis	II45
7.3.5	Uncertainties on the fit	II46
7.3.6	Limit setting	II47
8	Trigger Level Analysis	II49
8.0.1	Production of TLA data at triggers	II50
8.0.2	Data samples	II53
8.0.3	Calibration Procedure	II54
8.0.4	Event selection	II55
8.0.5	Background estimation	II56
8.0.6	Search phase results	II56
8.0.7	Limit setting	II57
8.0.8	Current status	II62
8.1	Improvements to the fitting procedure using Frequentist tools	II62
8.1.1	Setting limits with HistFitter	II63
8.1.2	Review and improvements of the fitting procedure	II64
8.2	Conclusions and outlook	II68

List of publications

In addition to the 214 ATLAS Collaboration publications, for which I qualified as author in June 2018, I have been the main author of the following publications:

- I **Impact of different compilers and build types on GEANT4 simulation execution time**

C. Marcon, O. Smirnova and S. Muralidharan

EPJ Web of Conferences 245, 05037 (2020) doi: 10.1051/epjconf/202024505037

- II **Studies of GEANT4 performance for different ATLAS detector geometries and code compilation methods**

C. Marcon, E. Elén, J.R. Madeira, B. Morgan, O. Smirnova and D. Smith

EPJ Web of Conferences 251, 03005 (2021) doi: 10.1051/epjconf/202125103005

All papers are reproduced with permission of their respective publishers.

Acknowledgements

This work was made possible by the passion and tireless dedication of many people at Lund University, at CERN and beyond. I would like to express my deepest gratitude to my main supervisor, Dr. Oxana Smirnova, for all the teachings and the opportunities received during these intense years. In addition, I also thank Dr. Caterina Doglioni for her advice and mentoring on the subjects of Dark Matter and TLA.

Popular summary in English

The Standard Model of particle physics (SM) took shape in the second half of the last century, merging together experimental and theoretical developments. It provides a consistent description of the elementary particles and of their interactions. However, based on cosmological and astrophysical observations, it has been found that the SM is only capable of describing a small portion of the mass and energy actually present in the universe. This large fraction of yet-undiscovered physics is commonly referred to as the physics Beyond the Standard Model (BSM). In fact, it is estimated that the SM can only cover 5% of the universe's content in terms of mass and energy, the remainder being attributed to non-luminous Dark Matter (DM) and Dark Energy.

Along with the refinement of the SM knowledge, the pursue of new DM discoveries is one of the hardest challenges of contemporary physics. Depending on the source of DM particles and on the interaction properties investigated by a given experiment, searches can be classified as *direct*, *indirect* or *collider* searches. The Large Hadron Collider (LHC) at CERN (Geneve, Switzerland), which has already played a fundamental role in the advances of the SM by paving the way to the discovery of the Higgs boson, is also designed to give a substantial contribution to the field of DM collider searches.

The most challenging aspect of DM production and detection at colliders is the inherent weakness of the expected signals. For the energy scales achievable at modern colliders, the typical DM cross-sections are much smaller than the competing SM ones. Therefore, to enhance the production of DM signals and to make new decay channels accessible, the planned experimental developments are mostly oriented towards a substantial increase of luminosity and center of mass energy. The direct result of this approach is a boost in the need for additional computing resources. Consequently, in addition to the extra load on trigger systems, Monte Carlo simulations will have to reproduce a much larger number of events in order to maintain an acceptable statistical power. Provided that the deployment of new and expensive hardware is currently not a viable option, efforts are ongoing to improve the efficiency of the existing codes in order to fully exploit the available computing power. In addition, statistical tools capable of analyzing the collected data in order to discover the DM signals require constant refinement: as limits are set on the highest possible strength of the searched signals, analysis codes need to be able to find weak localized excesses on top of the intense SM backgrounds.

This thesis offers a new perspective on the opportunities for optimizing the Geant4 Monte Carlo simulation workflows and presents an investigation of the behaviour of Frequentist statistical tools, as compared to Bayesian codes, applied to DM searches based on resonances appearing on the di-jet invariant mass spectrum.

Part I

Physics
at the Large Hadron Collider

Introduction to Part I

*[...]Wolfgang Pauli once asked me whether I had managed to actually understand Einstein's Theory of Relativity. I replied that I didn't really know the exact meaning of the word "understanding" in physics. I had no difficulties to understand the mathematical aspects of relativity, but I was not sure at all I had "understood" [...]. In fact, to be honest, I hardly grasped the whole matter, it looked incomprehensible to me [...]*¹.

As Heisenberg noted, already in 1921 the understanding of modern physics was not necessarily a straightforward endeavour. Today, just like one hundred years ago, the scientific community is devoting a vast effort to solve the great and fascinating mystery of Dark Matter, which is without doubt one of the most compelling open questions in contemporary physics.

This thesis has the objective of contributing to this large effort and deals with some of the challenges of today's physics at particle colliders. The first part is subdivided in three chapters covering the fundamental concepts of the Standard Model of particle physics (Chapter 1), the properties of the yet-undiscovered physics Beyond the Standard Model with a particular focus on Dark Matter searches (Chapter 2), and the setup of the ATLAS detector, one of the four major experiments currently operating at the Large Hadron Collider at CERN (Chapter 3).

¹English translation from [1].

Chapter 1

The Standard Model

Is it possible to identify something elementary, indivisible, from which all the objects surrounding us derive? And again, is there a single force from which all the interactions between objects around us derive?

Of course these questions are not only a need of the contemporary age, but since ancient times they have been an existential challenge for human beings: 25 centuries ago, Democritus of Abdera had already provided remarkable answers. Indeed, the philosopher speculated that by reducing matter into increasingly small fragments, it was possible to identify a fundamental entity that he called *atom* (uncuttable). Democritus then argued that these atoms, wandering in empty space and randomly interacting, give rise to the objects that surround us [2].

Today, the scientific understanding of the world is based on four fundamental forces: electromagnetism, strong and weak interactions and gravitation. The first three have been combined into a single theory: the *Standard Model* (SM) of particle physics that successfully describes the elementary particles and their interactions.

The Standard Model is the main topic of Sec. 1.1 of this introductory chapter. The attention will then be focused on a phenomenological description of hadron collider physics (Sec. 1.2).

1.1 Introduction to fundamental concepts

The Standard Model of particle physics provides a description of the fundamental components of matter and their interactions in a full relativistic quantum field theory [3]. The SM took shape mainly in the 1970s when important experimental and theoretical developments

merged together. It is called a "Model" but today it is one of the most complete theories in the history of science.

SM is a theory of interactions and describes the laws of nature by giving to each elementary particle¹ a susceptibility to certain forces. So far, it theorizes that all the visible matter is made up by fundamental constituents called *fermions*; two types of fermions can be distinguished: *leptons* and *quarks*.

All these particles interact with each other via four different fundamental forces: electromagnetic, weak, strong and gravity.

Unlike the electromagnetic, weak and strong interactions, the gravitational force is not incorporated in the SM². In this thesis, from now on, the gravity will not be considered further. It can be neglected since it is many orders of magnitude weaker than the other three forces [7].

In addition to fermions, the SM includes another category of particles: the *gauge bosons*, which transmit the forces acting as messengers.

In fact, an additional class of particles is needed to make a consistent theory of particle masses and interactions: the *scalar Higgs bosons*³.

A more detailed overview on the SM's particles and forces will be presented in Sec. 1.1.3.

1.1.1 Standard Model Lagrangian

In particle physics, it is conventional to describe the dynamics of a system in terms of the *Lagrangian function* or better the *Lagrangian density* [6].

The full formulation of the Standard Model Lagrangian would require several pages (see for example [8]) and it is beyond the scope of this thesis; a summarized version is presented in Eq. 1.1 [9]:

$$\begin{aligned}\mathcal{L}_{\text{SM}} = & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ & + \bar{\psi}\not{D}\psi + \text{h.c.} \\ & + \psi_i\gamma_{ij}\psi_j\phi + \text{h.c.} \\ & + |D_\mu\phi|^2 - V(\phi).\end{aligned}\tag{1.1}$$

¹Elementary particles are those without any substructure.

²The SM does not assign a gauge boson to the gravitational force. A mediator of this force, the *graviton*, has been proposed but has not yet been observed. The LIGO experiment, that confirmed the existence of gravitational waves, did not prove the existence of the graviton [4, 5]. Alternative theories to the Standard Model have provided some characteristics that this mediator should have: massless, without any charge and with spin equal to 2 [6].

³Currently SM includes only one Higgs boson experimentally detected at the CERN Large Hadron Collider in 2012, but more could exist [6].

- $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$: this term is the scalar product of the field strength tensor $F_{\mu\nu}$. It mathematically encodes all interacting particles except the Higgs boson and describes how they interact with each other.
- $+i\bar{\psi}\not{D}\psi + \text{h.c.}$: this term contains the description of the electromagnetic, weak, and strong interactions: it describes how bosons, except the Higgs, interact with fermions and anti-fermions. In this term the bosons' self-interactions are not considered. The *hermitian conjugate* (h.c.) of $+i\bar{\psi}\not{D}\psi$ ensures that the Lagrangian remains a real-valued function. Actually, since $+i\bar{\psi}\not{D}\psi$ is self-adjoint, the addition of h.c. is not required and is often omitted.
- $+\psi_i y_{ij} \psi_j \phi + \text{h.c.}$: this term describes how fermions couple to the Higgs field ψ and obtain mass; y_{ij} are the elements of the *Yukawa matrix* and represent the coupling parameters between particles and the Higgs field. In this case, the hermitian conjugate (h.c.) cannot be omitted since $+\psi_i y_{ij} \psi_j \phi$ is not self-adjoint. Indeed, it describes the interaction between a Higgs boson and fermions, whereas its hermitian conjugate describes the same interaction, but when anti-fermions are involved.
- $+|D_\mu\phi|^2$: this term describes how the weak gauge bosons couple to the Higgs field and thereby obtain their mass.
- $-V(\phi)$: this term describes the potential of the Higgs field and also describes how Higgs bosons couple to each other.

1.1.2 Symmetries

The SM was born as a theory of symmetries and has later evolved to include and explain violations of some of these symmetries. In the following section, attention is focused only on the mathematical foundation of the symmetric part of the SM.

An object is said to have a symmetry if there exist transformations which leave it unchanged. The concept of symmetry is crucial in particle physics: according to Noether's first theorem, every continuous symmetry of a physical system corresponds to a *conservation law* [10].

The most important examples of symmetries are related to very general features of space-time [11]. Indeed, the uniformity of space and time is reflected in the invariance of physical laws under translations of the spacetime coordinates:

- space translation invariance \Longleftrightarrow linear momentum conservation;
- time translation invariance \Longleftrightarrow energy conservation.

Another important feature is the isotropy of space which implies invariance under spatial rotations:

- space rotational invariance \iff angular momentum conservation.

Another example is the equivalence of coordinate systems moving relative to each other with constant velocity. This corresponds to the *Lorentz transformations*:

- Lorentz invariance \iff conservation of charge (C), parity (P) and time reversal (T) combined⁴ [11].

The symmetries mentioned above are inherent in all the fundamental interactions listed before.

In addition, other symmetries exist, not necessarily related to the spacetime properties, which occur only for a subset of phenomena. The presence or absence of symmetry characterizes the fundamental properties of particle interactions. Again, each of these symmetries is associated with an invariance under transformations in a given space characterized by internal degrees of freedom (e.g. quantum numbers which distinguish the particles involved). The theories in which the interactions are determined by the invariance under local transformations are called *gauge theories*.

All these transformations have a number of common features which allow them to be described within the same branch of mathematics: the *group theory* on which the treatment of symmetries is built [12]. Regarding group theory, only the definitions and properties that are necessary to understand the contents covered in this chapter will be outlined. For a more detailed discussion see, for example, [7, 11].

A set of transformations G is said to form a group if the group axioms are satisfied. In the group definition multiplication properties are also considered; in general, given two transformations g_1 and $g_2 \in G$, there is no requirement that $g_1 \cdot g_2 = g_2 \cdot g_1$, this means that multiplication commutativity is not a prerequisite for a group definition. Instead, it provides a means to make a distinction between *Abelian* and *non Abelian* symmetry groups: a group G is referred to as *Abelian* if multiplication of any two transformations in G is commutative and *non Abelian* otherwise.

In order to describe the particles and the SM interactions, three internal symmetries are needed: $U(1)$, $SU(2)$ and $SU(3)$.

$U(1)$ is the *Abelian* group of rotations in a plan. The letter U indicates that the group is unitary. In particle physics, $U(1)$ arises as the group of gauge transformations associated

⁴C, P and T transformations represent: the exchange of particles and anti-particles, the reflection of the three space axes and the reflection of the time axis, respectively.

with the conservation of electric charge (see Sec. 1.1.4).

In general, the Special Unitary group $SU(N)$ is the group of linear transformations of a complex N -dimensional vector space defined by unitary $N \times N$ matrices with determinant 1. The matrix multiplication is not in general commutative, therefore matrix groups are, as a rule, not Abelian. The simplest group of this series is $SU(2)$, consisting of complex 2×2 matrices. Since the dimension of $SU(N)$ is equal to $N^2 - 1$ [11], the dimension of $SU(2)$ is 3 as the number of the generators of its fundamental representation (i.e. the Pauli matrices). In particle physics, $SU(2)$ is related to the weak interactions (see Sec. 1.1.4). The symmetry group $SU(3)$ is associated to the strong interactions (see Sec. 1.1.4). The dimension of this group is 8 and the generators of its fundamental representation are the eight 3×3 Gell-Mann matrices.

1.1.3 Fermions

As anticipated in Sec. 1.1, the SM considers the existence of two different categories of particles: the *fermions* and the *bosons*.

Fermions are the basic particles of matter and are divided into *leptons* (l) and *quarks* (q). Both have spin $\frac{1}{2}$, are subject to the *Pauli exclusion principle* (which states that two or more identical fermions cannot occupy the same quantum state) and obey to the *Fermi-Dirac statistics*⁵. So far, six different leptons are known, arranged into three *families* (or *generations*) with increasing masses as summarized in Tab. 1.1. Each family includes two particles of different flavor: one massive and with unitary electric charge (e^- , μ^- and τ^- ; *electron*, *muon* and *tauon*, respectively) and one electrically neutral called *neutrino* (ν_e , ν_μ and ν_τ) with non vanishing mass⁶. A *lepton number* (L) can be defined for each family. It is a conserved additive quantum number: a particle reaction can only occur if the sum of the L values in the initial state equals that in the final state [12]. The theoretical reasons why it should be absolutely conserved are not known: the conservation of this quantity is not supported by a fundamental symmetry but it is only assumed because no violations have yet been observed.

Similarly, there are six flavors of *quarks* (see Tab. 1.1). The existence of quarks was theorized independently and almost simultaneously by M. Gell-Mann and G. Zweig in 1964 [16]. Gell-Mann called these particles quarks, using a word that had been introduced by James Joyce in his novel *Finnegans Wake* [16]. The experimental evidence of their existence began

⁵The statistics obeyed by a particle determines how the wave function ψ , describing an ensemble of identical particles, behaves under exchange of any pair of particles.

Under exchange of identical bosons: $\psi \Rightarrow +\psi$ (ψ is symmetric).

Under exchange of identical fermions: $\psi \Rightarrow -\psi$ (ψ is anti-symmetric) [13].

⁶The SM does not require neutrinos to have mass. However, the observation of neutrino oscillations [14] indicates that these particles do have a non-zero mass (the current experimental upper bound is 0.2 eV [15]). Having mass allows neutrinos to oscillate to other families.

in the late 60s of the last century when the *up*, *down* and *strange* quarks were discovered thanks to deep inelastic scattering experiments (at SLAC). This experimental proof was then concluded in 1995 with the discovery of the *t* quark at Fermilab [17, 18]. This quark appears to be, so far, the heaviest fundamental particle ever observed. For historical reasons they are called: *up*, *down*, *strange*, *charm*, *bottom* and *top* and they are denoted by the first letters of their names [6]. As presented in Tab. 1.1, they fall into doublets also called *families* or *generations*. They carry non-integer electric charge: *u*, *c* and *t* carry $+\frac{2}{3}e$, while *d*, *s* and *b* $-\frac{1}{3}e$, where *e* is the magnitude of the electron's electric charge. If the history of the discovery had been different, the electric charge of the electron would probably have been defined equal to $-3e$.

All quarks carry quantum number called *baryon number* (*B*); it is equal to $\frac{1}{3}$ for each quark. Like the lepton number, conservation of *B* is experimentally verified with a very good accuracy; although no theoretical reason is known why it should be conserved.

Quarks, in addition to the electric and weak interactions, are also sensitive to the strong force as carriers of the *color charge* that comes in three different variations: red (*r*), green (*g*) and blue (*b*)⁷. Quarks do not exist in nature as free particles but they combine to make color-neutral bound states. Particles built up from quarks are called *hadrons*, in particular they are named *mesons* if formed by two quarks, and *baryons* when three quarks are involved⁸. In this context it is necessary to make a clarification: in the distinction between *mesons* and *baryons*, only the *valence quarks* have been considered. They define the hadronic macroscopic properties, such as charge and spin. The valence quarks are actually immersed in a sea of virtual quarks. The latter are also called *sea quarks* and are produced by *gluons* (strong force gauge bosons) which, due to quantum fluctuations, split into quark-antiquark pairs which then annihilate back to gluons. Gluons, sea and valence quarks are collectively called *partons*.

Despite the large number of leptons and quarks existing in nature, the matter that surrounds us is made up only of fermions and quarks belonging to the first family. The other quarks also combine and give rise to heavy hadrons which, however, are unstable and rapidly decay ($\sim 10^{-13}$ s) into *u* and *d* combinations. The heaviest leptons μ and τ also undergo rapid decay into neutrinos. These heavy and unstable composite particles can only be observed in high energy collisions at accelerators or naturally in cosmic rays [13].

⁷This color is not related to the colors we see.

⁸The term baryon comes from the Greek " $\beta\alpha\rho\nu\varsigma$ " (barys) which means heavy. In fact, baryons with more than three quarks have also been observed (e.g. *tetraquarks* and *pentaquarks* composed of 4 and 5 quarks respectively). Theorized since the beginning of the quark model, they have been observed experimentally only in the 2000s [19, 20].

Table 1.1: Overview of the fermionic sector of the Standard Model. The subscripts LH and RH stand for Left-Handed and Right-Handed, respectively. Quantum numbers I , I_3 , Y and l are defined in the text.

Fermions												
	1 st generation		2 nd generation		3 rd generation		$Q(e)$	I	I_3	Y	B	Spin
	Flavor	Mass	Flavor	Mass	Flavor	Mass						
Leptons	$\begin{pmatrix} e \\ \nu_e \end{pmatrix}_{\text{LH}}$	$\sim 0.511 \text{ MeV}$	$\begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_{\text{LH}}$	$\sim 105.6 \text{ MeV}$	$\begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_{\text{LH}}$	$\sim 1.777 \text{ GeV}$	-1	$\frac{1}{2}$	$-\frac{1}{2}$	-1	1	$\frac{1}{2}$
		$< 2 \text{ eV}$		$< 2 \text{ eV}$		$< 2 \text{ eV}$	0	$\frac{1}{2}$	$+\frac{1}{2}$	-1	1	$\frac{1}{2}$
	e_{RH}	$\sim 0.511 \text{ MeV}$	μ_{RH}	$\sim 105.6 \text{ MeV}$	τ_{RH}	$\sim 1.777 \text{ GeV}$	-1	0	0	-2	1	$\frac{1}{2}$
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_{\text{LH}}$	$\sim 2.16 \text{ MeV}$	$\begin{pmatrix} c \\ s \end{pmatrix}_{\text{LH}}$	$\sim 1.27 \text{ GeV}$	$\begin{pmatrix} t \\ b \end{pmatrix}_{\text{LH}}$	$\sim 172.9 \text{ GeV}$	$\frac{2}{3}$	$\frac{1}{2}$	$+\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{1}{2}$
		$\sim 4.67 \text{ MeV}$		$\sim 93 \text{ MeV}$		$\sim 4.18 \text{ GeV}$	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{1}{2}$
	u_{RH}	$\sim 2.16 \text{ MeV}$	c_{RH}	$\sim 1.27 \text{ GeV}$	t_{RH}	$\sim 172.9 \text{ GeV}$	$\frac{2}{3}$	0	0	$\frac{4}{3}$	0	$\frac{1}{2}$
	d_{RH}	$\sim 4.67 \text{ MeV}$	s_{RH}	$\sim 93 \text{ MeV}$	b_{RH}	$\sim 4.18 \text{ GeV}$	$-\frac{1}{3}$	0	0	$-\frac{2}{3}$	0	$\frac{1}{2}$

In 1931, P. Dirac predicted the existence of *antiparticles*: objects with the same mass, lifetime and spin as the corresponding particles but with opposite sign of charges (both electric and color), leptonic and baryonic numbers and magnetic momentum [6, 13].

The existence of antiparticles is a general property of both fermions and bosons, but only for the first type of particles the difference in the number of fermions and anti-fermions is a constant (the leptonic and baryonic numbers are conserved). Some neutral particles, such as photons or the π^0 meson, are coincident with their own antiparticles⁹. The standard notation for antiparticles is an overbar but, in some cases, it is also possible to specify the charge instead (e.g. positron is denoted as e^+).

1.1.3.1 Left-handed and right-handed particles

Helicity (H) is the quantity that measures the sign of a particle's spin projection along the direction of the motion. As shown in Fig. 1.1, $H = -1$ corresponds to a *left-handed* (LH) screw, whereas, particles with $H = +1$ are *right-handed* (RH) [13].

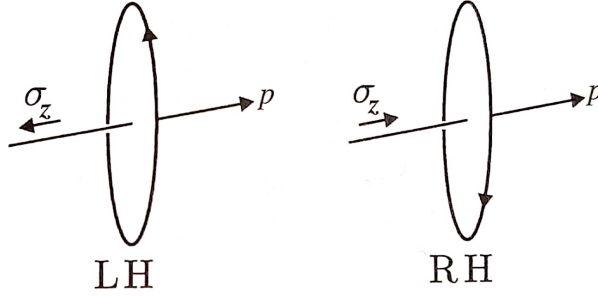


Figure 1.1: Representation of left-handed (LH) and right-handed (RH) particles [13].

Helicity is not a Lorentz-invariant quantity since it depends on the reference frame¹⁰. Indeed, given a state of helicity, it is always possible to find a Lorentz transformation capable of changing the sign of the helicity unless the particle has zero mass. If the particle is massless, it travels at the speed of light c , thus it is impossible to reverse the helicity with a Lorentz transformation to a second reference frame which relative velocity is $v < c$. As presented in Table 1.1, only left-handed neutrinos are considered in the SM and there is no right-handed counterpart. This can explain why, according to SM predictions, neutrinos are expected to be massless. *Helicity*, as well as *chirality*, are of crucial importance for the

⁹For neutrinos, however, it has not yet been possible to prove whether ν is equivalent to $\bar{\nu}$. If this were true, then it would mean that the neutrino is a *Majorana particle*. Otherwise it would be a *Dirac particle*.

¹⁰A quantity related to helicity but frame independent is *chirality* (see for example [3] for a more detailed explanation).

charged weak interactions (see Sec. 1.1.4.2).

1.1.4 The SM forces

As anticipated, the SM also explains the interactions that occur between fermions. These interactions are described in terms of the exchange of mediators called *gauge bosons*. These messenger particles have an integer spin, do not obey Pauli’s exclusion principle and behave according to the *Bose-Einstein* statistics [7]. In the following, the attention will be focused on each of the three forces incorporated in the Standard Model.

1.1.4.1 Electromagnetic force

Electromagnetic interactions are responsible for all extra-nuclear phenomena (e.g. bound states of electrons with nuclei and intermolecular forces in liquids and solids). The physical theory that describes the electromagnetic force is called *quantum electrodynamics* (QED). Its classical formulation is to be attributed to J. C. Maxwell (1831-1879), while the quantum version required a joint effort by numerous physicists, reaching a definitive form thanks to S. Tomonaga, R. Feynman and J. Schwinger in the 1940s [21].

These interactions are mediated by a boson called *photon* (see Tab. 1.2). Photons are massless and do not have electric charge; they can couple to all electrically charged particles. This type of interaction obeys the gauge symmetry group $U(1)$ (see Sec. 1.1.2). As discussed in Sec. 1.1.2, the presence of symmetries implies the conservation of some quantity: in QED the electric charge is conserved (in addition to the lepton and baryon numbers). Furthermore $U(1)$ is an Abelian group of dimension 1. The first feature leads to the absence of reciprocal interaction between photons; the second reflects the fact that there is only one type of photon (i.e. all photons are indistinguishable from each other). From a phenomenological perspective, both these properties are confirmed for the electromagnetic force.

1.1.4.2 Weak force

Weak interactions affect all fermions including neutrinos, which are not sensitive to any other force. These interactions play a crucial role in the energy generation process of the Sun, in the formation of heavy elements and in β -decays [13]. They were completely unknown to classical physics and, from the beginning, their theoretical description was formulated according to relativistic quantum mechanics. This interaction is called “weak” because, in general, it is less intense than both the electromagnetic and the strong force.

E. Fermi proposed a first formulation of the weak interaction in 1933 [22]; this proposal was then further developed by T. D. Lee, C. N. Yang, R. Feynman, M. Gell-Mann and many others in the 1950s. The current formulation is due to S. Glashow, S. Weinberg and A. Salam: in the 1960s, they explained the parity violation that occurs in these interactions and also unified the electromagnetic and weak forces unto a single force called *electro-weak* [23, 24].

The gauge symmetry of the weak force is the non-Abelian $SU(2)$ group; this means that there are three different gauge bosons that act as mediators, W^- , W^+ and Z^0 and that they can interact among themselves (see Table 1.2). The *weak isospin* I is the conserved charge of the weak interaction. Unlike right-handed fermions, which have null I and I_3 ¹¹, left-handed fermions are organized into doublets with a weak isospin equal to $\frac{1}{2}$ and I_3 equal to $\pm\frac{1}{2}$. W^+ and W^- have a unitary I and I_3 equal to $\pm\frac{1}{2}$, respectively; Z^0 , on the contrary, has null I and I_3 .

W bosons have mass and are electrically charged and this causes the electric charge of the fermions interacting with them to increase (or decrease) by one unit.

The *charged weak interactions* are the only ones allowing the change of flavor. This statement comes with a caveat: lepton flavor is conserved since the leptonic charged weak interactions connect members of the same generation (e.g. an e^- is converted into ν_e with the emission of W^- but the e^- never goes to μ^- or ν_μ); instead, this behavior is not completely confirmed when quarks are involved. In general, they tend to mimic the leptons behavior. Nevertheless, in order to explain, for example, the decay of particles such as Ω^- or Λ ¹², it is necessary to assume the conversion of a s quark into an u quark. Thus, with quarks, the flavor is not always conserved [7].

Another important peculiarity of the charged weak interactions is represented by the parity violation: it is experimentally observed that W bosons only couple to fermions with left-handed helicity and to right-handed antiparticles. On the contrary, the boson Z_0 can interact both with left-handed and right-handed fermions. It is massive, as W , but it does not carry electric charge. Unlike charged weak interactions, which were considered from the start, the existence of *neutral weak* processes was not considered until 1958; their presence was experimentally confirmed only in 1973 at CERN with bubble chamber *Gargamelle*.

As already mentioned, in the late 1960s, S. Glashow, S. Weinberg and A. Salam proposed the *electroweak theory* in which electromagnetic and weak interactions are treated as different aspects of a single interaction [25]. This theory postulates the existence of four massless bosons arranged in a triplet and a singlet: $W_\mu^{(1)}$, $W_\mu^{(2)}$ and $W_\mu^{(3)}$ and a B_μ respectively¹³. $W_\mu^{(1)}$, $W_\mu^{(2)}$ and $W_\mu^{(3)}$ are the components of $I = 1$ weak isospin triplet of the group $SU(2)$

¹¹ I_3 is the isospin's third component.

¹² $\Omega^- \rightarrow \Lambda + K^-$ and $\Lambda \rightarrow p^+ + \pi^-$ [7].

¹³Superscripts 1, 2, 3 indicate the components in "isospin space". The subscript μ represents the space-time components of a 4-vector [13].

while B_μ has $I = 0$ and belongs to the group $U(1)$. The symmetry group of the electroweak force is therefore $SU(2) \times U(1)$ and the conserved quantity is the *weak hypercharge* defined as: $Y = 2(Q - I_3)$ where Q is the electric charge.

At this point, however, there seems to be an unsolvable contradiction: the experimentally observed W bosons have a non-negligible mass, while the electro-weak theory assumes the existence of massless particles.

To give bosons a mass, a process called *spontaneous symmetry breaking* has been incorporated into the electroweak theory. This mechanism was formulated by P. Higgs [26], R. Brout and F. Englert [27]. It introduces a new spin-zero quantum field, called *Higgs field*, which permeates the universe. The gauge bosons and fermions can interact with the Higgs field and, in its presence, they are no longer massless¹⁴. The interactions with this field causes a spontaneous breaking of the electro-weak symmetry group $SU(2) \times U(1)$. This break ensures that the four electro-weak bosons $W_\mu^{(1)}$, $W_\mu^{(2)}$ and $W_\mu^{(3)}$ and B_μ mix to give the experimentally observed bosons:

$$Z^0 = \cos \theta_W W_3 - \sin \theta_W B \quad (1.2)$$

$$\gamma = \sin \theta_W W_3 + \cos \theta_W B \quad (1.3)$$

$$W^+ = \frac{1}{\sqrt{2}}(W_1 - iW_2) \quad (1.4)$$

$$W^- = \frac{1}{\sqrt{2}}(W_1 + iW_2) \quad (1.5)$$

where θ_W is the *weak mixing angle*, also called *Weinberg angle*. It is an arbitrary constant which value has been established experimentally.

A detailed review of the physics of the Higgs boson is outside of the scope of the present thesis. Further discussion on this subject can be found for example at Ref. [6].

1.1.4.3 Strong force

The theory of the strong force is called *quantum chromodynamics* (QCD). Strong interactions are only felt by *quarks*. These forces are responsible for binding quarks together into hadrons and, indirectly, for holding the neutrons and protons together in nuclei [13]. In QCD, color charge plays the same role that electric charge and isospin have in QED and weak interactions, respectively. The strong force obeys the non-Abelian symmetry group $SU(3)$. This implies the existence of eight different color charge carriers called *gluons* (see

¹⁴Fermions acquire mass through the *Yukawa coupling*. The size of this coupling is measured only experimentally and is currently not explained by theory.

Table 1.2) that can undergo self interactions. The gluons are massless, they are not affected by either the electromagnetic or the weak forces but are carriers of two color charges: a charge and an anticharge. This peculiarity has consequences that are far from negligible: the strong coupling constant α_s ¹⁵ depends on the scale at which it is measured. Indeed, α_s shows a dependence on the distance and on the energy at which it is probed: at small distances¹⁶ (i.e. high energies), the self-coupling of gluons leads to “anti-screening” effect and this results in a weakening of α_s [7]. This effect is called *asymptotic freedom* because the quarks behave as almost free. At this regime, QCD can be treated using *perturbative calculations*¹⁷.

On the contrary, α_s becomes larger and larger for increasing distances (i.e. low energies), causing a *confinement* effect inside the hadrons. When trying to separate two quarks, the energy required increases with distance until a threshold energy is reached for which it is energetically more favorable to create a new quark pair rather than to continue separating the original couple. This new pair forms a new colorless state with the original quarks giving rise to a process called *hadronization* (a more detailed discussion on this will be provided in Sec. 1.2). This explains why quarks are bound in color-neutral states and cannot be observed free. The dependence of α_s on the energy¹⁸ at which it is probed can be expressed, at leading order, as follows (Eq. 1.6):

$$\alpha_s(Q) = \frac{12\pi}{(33 - 2n_f) \log(\frac{Q^2}{\Lambda^2})} \quad (1.6)$$

where n_f represents the number of quark flavors relevant at the energy scale considered ($Q^2 > q_m^2$, where q_m represents the quark mass) and Λ is the energy scale at which the perturbative coupling would nominally become infinite. This Landau pole is experimentally determined to be around 250 MeV. The coupling strength tends to zero for Q values that go to infinity: this is the *asymptotic freedom* regime.

For small values of Q , α_s tends to increase, becoming even larger than 1 for $Q \lesssim 100$ MeV. At these energies, confinement dominates and it is no longer possible to treat QCD with perturbation theory. In Fig. 1.2 the strong coupling constant α_s as a function of the momentum transfer Q is shown [28].

This strong energy dependence of α_s , called *running coupling constant*, leads to a complex phenomenology that will be investigated in Sec. 1.2.

¹⁵ α_s measures the strength of the strong force.

¹⁶Smaller than the size of a hadron.

¹⁷Perturbative calculations, similarly to Taylor expansions, provide an approximation of complex calculation by means of a series of terms of increasingly higher accuracy [11].

¹⁸The energy, in this case, is parameterized as moment transferred Q .

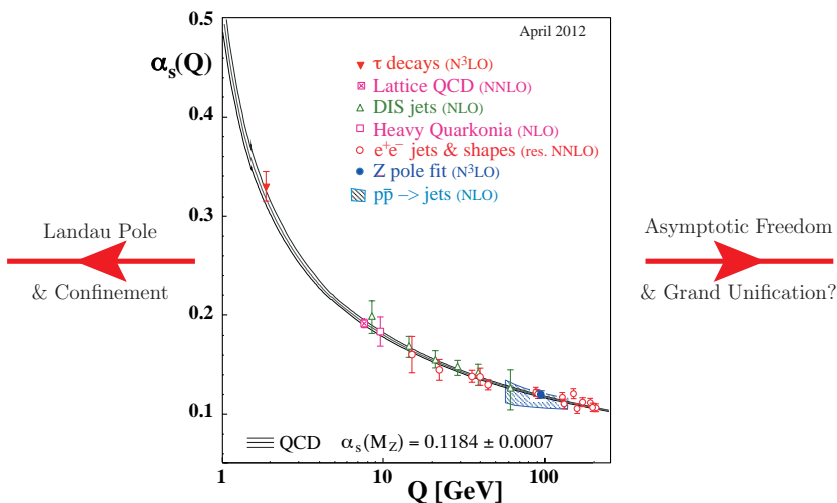


Figure 1.2: Illustration of the running of α_s in a theoretical calculation (band) and in physical processes at different characteristic scales [28].

In the mid 1970s H. Georgi and S. Glashow suggested that the observed gauge symmetries of the SM could be accommodated within a larger $SU(5)$ group. In this *Grand Unified Theory*, the coupling constants of the SM are found to converge at an energy scale of about 10^5 GeV [3].

Table 1.2: Summary of the most relevant features of the Standard Model forces and their mediators.

Forces and gauge bosons					
Force	Acts on	Mediator			
		Name	Mass	$Q(e)$	Spin
Electromagnetic	All electrically charged particles	Photon (γ)	0	0	1
Weak	Quarks, leptons, electroweak gauge bosons	Z^0	~ 91.2 GeV	0	1
		W^+	~ 80.4 GeV	+1	1
		W^-	~ 80.4 GeV	-1	1
Strong	Quarks and gluons	Eight gluons (g)	0	0	1

1.2 Phenomenology of QCD

The peculiar properties of QCD together with the compositeness of hadrons make the description of hadronic collisions extremely complicated.

In this section the phenomenology of a hadronic collision, specifically proton-proton, will be described and the fundamental concepts for the calculation of *hadronic cross sections* will be introduced.

1.2.1 Evolution of a hadronic collision

A *hadronic collision* begins with two hadrons that are accelerated and collide at relativistic energies and ends with showers of mesons and baryons characterized by much lower energies. The complete description of these interactions, outlined in Fig. 1.3, can be organized in different sub-processes which will be examined following their chronological development. The proton, as already explained, is not an elementary particle as it appears to be composed of partons. The partons include *valence quarks* (u , u , d), *sea quarks* and *gluons*.

In the scheme, the two protons involved in the interaction are indicated with the large green ellipses and the three green lines represent the three valence quarks. In a p-p collision, at TeV energy scale, the initial hard scattering process does not involve the proton as a whole but instead it occurs between two of the partons. This hard interaction is represented by the big red circle.

In general, gluons can undergo a splitting process ($g \rightarrow gg$, $g \rightarrow q\bar{q}$) or a gluon radiation can happen ($q \rightarrow qg$ or $\bar{q} \rightarrow \bar{q}g$); depending on when these processes occur, (i.e. before or after the hard scattering event) they are referred to as *initial state radiation* (ISR) and *final state radiation*, (FSR) respectively.

Focusing the attention on the FSR, the partons involved in the hard-scatter interactions are violently accelerated and, similarly to the bremsstrahlung radiation emitted by accelerated electric charges, also the accelerated color charges can emit gluons, thus depositing energy. These high energy final state partons can emit further QCD radiation and generate the so-called *parton showers*. In principle, the formation of parton showers could be treated from a perturbative perspective by considering higher-order corrections to the hard-scatter interactions. However, in practice, an approach based on successive approximations is adopted and only the dominant contributions are considered (i.e. *parton splitting* and *soft gluon emission*).

In Fig. 1.3 the parton showers and their evolution are indicated with red lines, while the small red circles represent the products of hard interactions (in this specific case two top quarks and a Higgs boson). The showering process leads partons to loose energy and when the energy is such that confinement becomes dominant ($Q < 1$ GeV), the showering stops

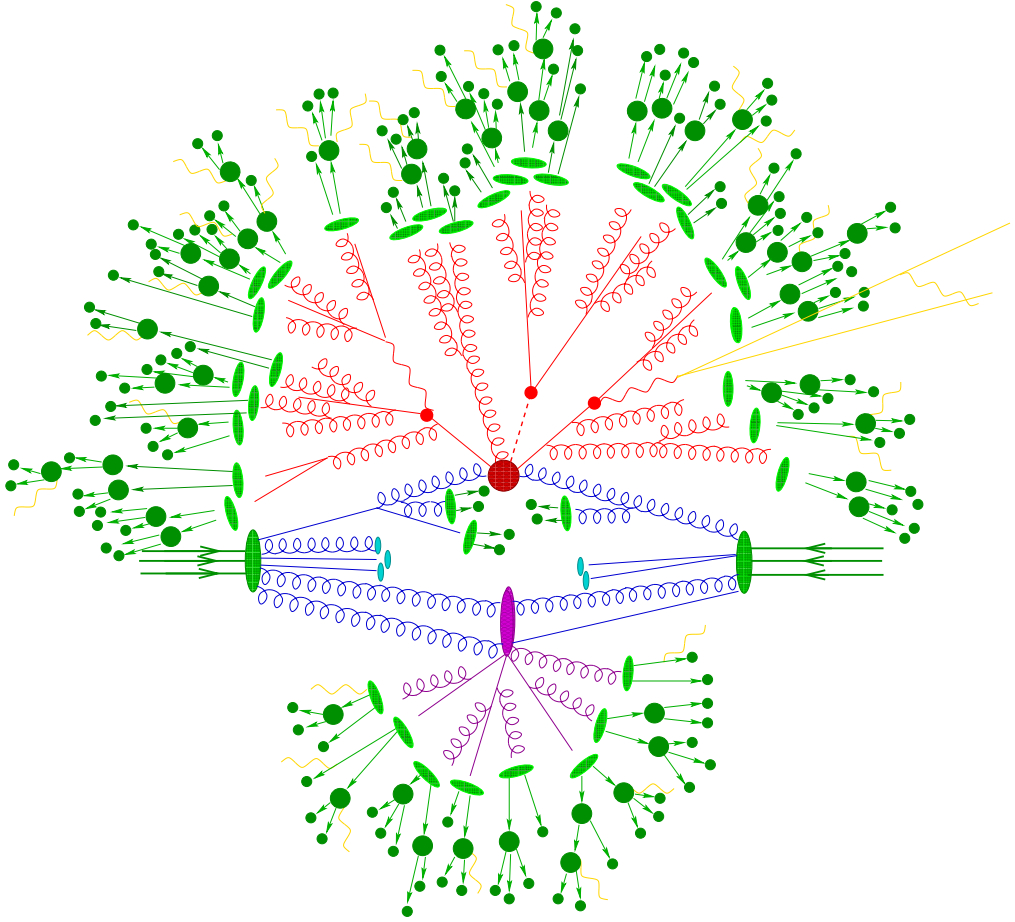


Figure 1.3: Sketch of a $pp \rightarrow ttH$ event. Two partons from the colliding protons (large green ellipses) emit *initial state radiation* and then interact in the *hard process* (central big red blob) giving rise a *parton shower* (red). The resulting partons *hadronize* into colorless states (light green ellipses) then decaying into stable particles (green circles).

The purple blob, in the lower hemisphere, represents a *secondary interaction* between proton remnants. Also in this case, a parton shower is created (purple). This secondary interaction, together with the beam remnants (light blue ellipses), represents an *underlying event*. Yellow lines signal soft photon radiation that can occur at any stage [29].

and the resulting partons can hadronize into colourless states: the *hadronization* process begins. In this phase, starting from independent partons, baryons and mesons are formed. The latter are indicated in Fig. 1.3 with the light green blobs.

The hadronization process cannot be treated with the perturbative QCD approach because the strong coupling α_s is too large. Until now no exact description has been developed but several models allowing to make predictions on the hadron formation are available. Two different approaches will be briefly presented below which underlie different models often used to simulate the hadronization process: *string* and *cluster* hadronization.

In *cluster models* ([30]), the gluons produced in the parton shower are split into $q\bar{q}$ pairs. These combine to give rise to colorless clusters which can then decay into lighter clusters of hadrons pairs depending on the available phase-space (Fig. 1.4).

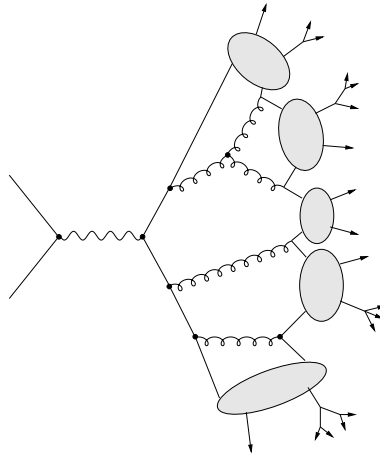


Figure 1.4: Cluster hadronization model [31].

In general, *string models* (Fig. 1.5) lay their foundations on the dynamics of the colour flux between the $q\bar{q}$ pairs. As already pointed out, gluons experience self-interactions and this causes the color field lines between $q\bar{q}$ to be compressed into a string-like region where the color field is strong.

According to the *Lund model* [32], one of the existing string models, the force between two moving quarks is treated as a classical string: as the distance increases, the kinetic energy of the quarks is transformed into potential energy of the string. As presented in Fig. 1.6, when this energy exceeds the mass of $q\bar{q}$ pair, it becomes more convenient to break the original string into two new strings with a q at one end and an \bar{q} at the other. The new strings then stretch and break into more quark pairs. This process continues until all the potential energy is converted to quarks which are connected to each other by small gluon strings. The latter can be regarded as hadrons and the hadron type depends on the flavor of the quark pairs created in the breaking.

These primary hadrons eventually decay into stable particles that, in Fig. 1.3, are represented

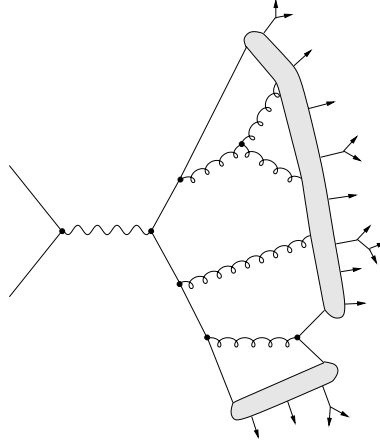


Figure 1.5: String hadronization model [31].

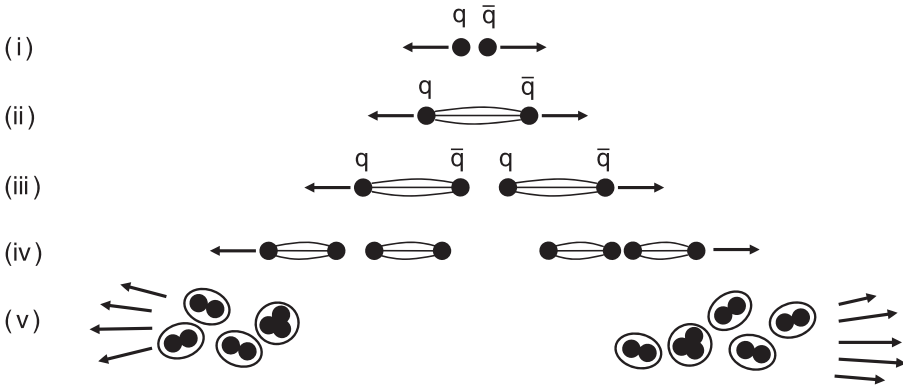


Figure 1.6: Qualitative picture of the steps in the hadronization process [3].

with the green circles. These stable particles are the only experimentally observable objects and for this reason they are used as a signature for studying QCD processes.

Since both showering and hadronization conserve momentum, if a parton has a sufficiently large momentum, then also the particles it produces will maintain a significant amount of their momenta in the same direction. This leads to the formation of a localized spray of hadrons which is called *jet*. By measuring the properties of the jets it is possible to reconstruct the features of the original parton which cannot be observed directly.

So far the attention has been focused on the most energetic part of the parton interactions, the hard-scattering. In fact, in a high energy p-p collision, also secondary hadronic activities can occur. They are typically much softer than the primary ones and are called *underlying events*. This type of events are represented in the lower hemisphere of Fig. 1.3.

Part of these secondary activities come from the remnant partons (purple ellipse) that do not directly participate in the hard process. There is a gluon field among these particles and a parton shower is produced (purple lines). As in the hard scattering process, the showering is followed by the hadronization step and then particles decay into stable objects. The beam remnants (light blue ellipses) and the contributions from multiple parton interactions are also part of this underlying event category. Multiple interactions happen when more than one parton of each colliding proton is involved in the interaction. The probability of these additional interactions depends on the *impact parameter*, which gives an estimation of the overlap between the protons in the initial collision.

Some yellow traces are also drawn in Fig. 1.3: they represent the electromagnetic radiation that is emitted by the charged particles.

1.2.2 Jets

As discussed in the previous section, the hadronization results in a spray of hadrons called *jet*. In the 50s, jets began to be observed in cosmic rays events. The term was applied to sprays of particles emerging from the collisions of energetic cosmic rays in an emulsion target [33]. The non-isotropic distribution of hadrons in their final state was then also tested by e^+e^- colliders, such as the JADE experiment at PETRA, starting from the 1970s [15].

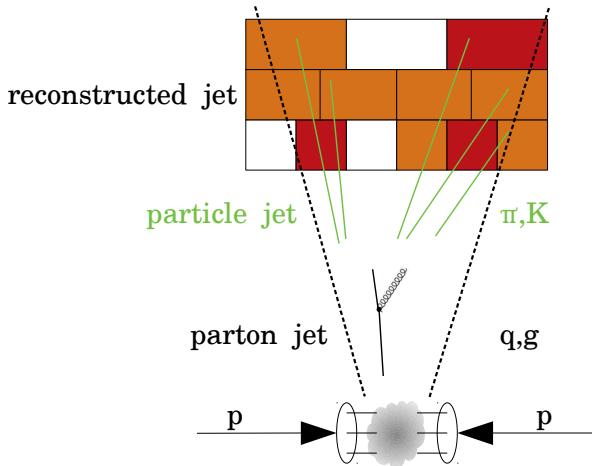


Figure 1.7: The sketch shows the formation of a *particle jets*. After the hard scatter, a parton showers and, eventually, hadronizes forming a collimated and approximately cone-shaped jet of stable particles that can be experimentally detected.

The hadronization of these partons results in a collimated cluster of particles, the so-called

particle-jets, that reach the detector, where energies and momenta are measured (Fig. 1.7). The identification and match of a jet object with the original parton that might have originated such a jet, occur through the application of *clustering algorithms*. These algorithms define the procedure by which final-state observables are clustered into a single jet object. These algorithms are of fundamental importance for many analyses; for this reason stability and efficiency are primary requirements [34]. These details will be examined in depth in Chapter 7.

1.2.3 The PDFs

When two protons collide at TeV energy scale, the interactions are not between the protons as a whole but, on the contrary, involve their constituents, both gluons and quarks. Furthermore, due to asymptotic freedom, the partons are not strongly bound within the proton and the overall momentum is given by the net effect of their unrestricted motion. The *parton distribution functions* (PDFs) are able to describe the initial state of the protons, before the collision, in terms of their partons. These functions give the probability to find a parton with a certain momentum fraction x ¹⁹ at a given energy scale Q .

The QCD theory is not able to predict the proton's structure so it is not possible to deduce the PDFs from first principles, instead they are experimentally determined by deep inelastic scattering experiments [36].

Two examples at different Q^2 are presented in Fig. 1.8. Based on these plots, it is possible to conclude that the PDFs of valence quarks tend to dominate for high values of x : if the parton carries more than 20% of the longitudinal momentum, it is very likely to be an up or a down quark or a gluon. For higher Q values, the contributions at low x become larger; this effect is due to the increased number of gluons and sea quarks of all flavors. Considering the left plot of Fig. 1.8, it is also possible to note that the number of accessible sea quarks depends on the energy scale at which the proton is probed. These considerations lead to the conclusion that the fraction of the proton momentum carried by sea quarks and gluons depends on the energy transfer in the collision.

PDFs cannot be calculated from first principles; by means of the *Dokshitzer-Gribov-Lipatov-Altarelli-Parisi equations* (DGLAP) (Eq. 1.7) it is however possible to infer a PDF for a given Q' starting from the measurement of the same PDF for $Q \neq Q'$ [36, 38].

¹⁹ x is the Björken variable. It is a Lorentz invariant dimensionless quantity that can be interpreted as the fraction of the proton momentum that each parton carries [35].

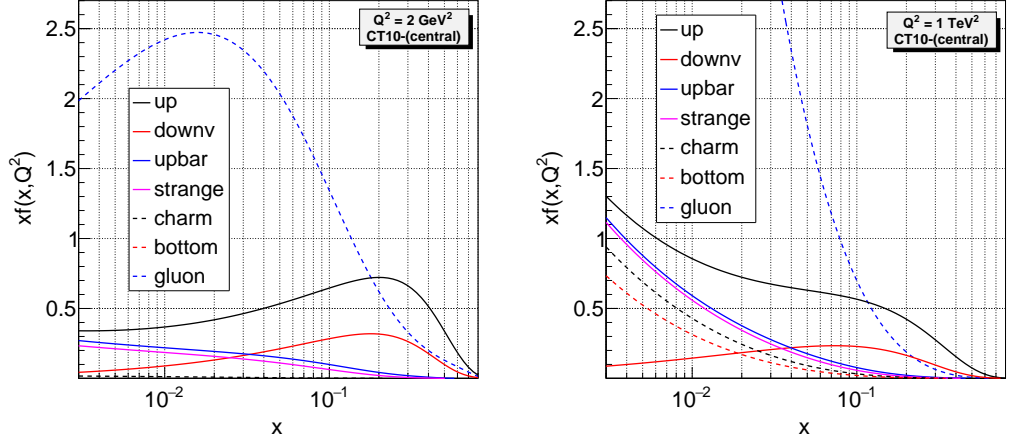


Figure 1.8: Parton distribution functions showing the distribution of momentum between q , \bar{q} and g . In the left plot the PDFs are evaluated at $Q^2 = 2 \text{ GeV}^2$ while in the right plot PDFs are evaluated at $Q^2 = 1 \text{ TeV}^2$. The PDFs for q and \bar{q} of non-valence quark flavors are assumed to be identical [37].

$$\begin{aligned} \frac{\partial q_i(x, Q^2)}{\partial \log Q^2} &= \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} \left\{ P_{q_i q_j}(z, \alpha_s) q_j\left(\frac{x}{z}, Q^2\right) + P_{q_i g}(z, \alpha_s) g\left(\frac{x}{z}, Q^2\right) \right\} \\ \frac{\partial g(x, Q^2)}{\partial \log Q^2} &= \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} \left\{ P_{g q_j}(z, \alpha_s) q_j\left(\frac{x}{z}, Q^2\right) + P_{g g}(z, \alpha_s) g\left(\frac{x}{z}, Q^2\right) \right\} \end{aligned} \quad (\text{I.7})$$

where $g(x, Q^2)$ and $q_i(x, Q^2)$ represent the gluon and the quark PDF, respectively. The four $P_{ab}(z, \alpha_s)$ functions are known as *splitting functions* and describe the probability distribution for a parton to radiate another parton [36].

1.2.4 Factorization theorem for hadronic cross sections

In a particle interaction process, the *cross section* quantifies the probability for an initial state to evolve into a new state. It has the unit of an area, and it is measured in barn (b): $1 \text{ b} = 10^{-28} \text{ m}^2$. In general, in particle physics, the calculations of transition rates, at the level of fundamental particles, are based on the *Fermi's golden rule* [3]. This principle states that the rate for a process is given by the square modulus of the quantum mechanical amplitude integrated over the available phase space.

The computation of a cross-section for a hadronic process requires further elaboration since it contains both perturbative and non perturbative behavior simultaneously. Perturbative

calculations can indeed be used for treating hard-scattering interactions, but the processes taking place before and after these high-energy collisions might occur at much lower scales and must therefore be treated in a different way. These two regimes, perturbative and non-perturbative, are independent of each other so they are factorizable. The *factorization theorem* [39] states that the hadronic cross-section $\sigma(P_1, P_2)$ between two hadrons (protons in this case), can be computed by weighting the partonic sub-process $\hat{\sigma}_{ij}(\mu_R^2, \mu_F^2)$ with the PDFs f_i, f_j and then integrating over the possible combinations (Eq. 1.8):

$$\sigma(P_1, P_2) = \sum_{i,j} \int dx_1 f_i(x_1, \mu_F^2) \int dx_2 f_j(x_2, \mu_F^2) \hat{\sigma}_{ij}(\mu_R^2, \mu_F^2) \quad (1.8)$$

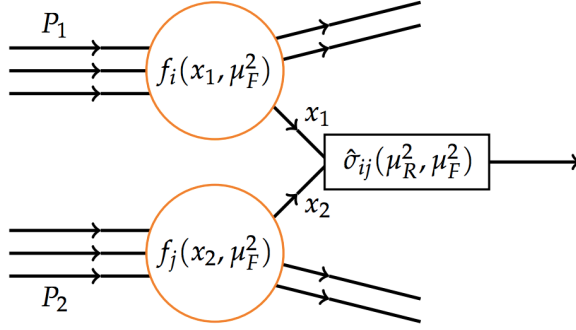


Figure 1.9: Schematic representation of the factorizable process in a pp collision.

P_1 and P_2 are the total momenta of the two colliding protons, $f_{i,j}$ are the PDFs and x_1, x_2 are the *Bjorken x -values* of the two partons. As discussed, x_1 and x_2 can be interpreted as the fraction of the proton momentum each parton carries. The summation is over all types of partons in the two colliding hadrons.

This factorization, outlined in Fig. 1.9, introduces a *factorization scale* μ_F^2 ²⁰ which marks the transition between long and short distance physics (i.e. the transition between the non perturbative and the perturbative regime).

The partonic subprocess $\hat{\sigma}_{ij}(\mu_R^2, \mu_F^2)$ depends also on the renormalization scale for the strong running coupling constant μ_R even though the cross section should in principle be independent of both μ_R and μ_F . This can only occur if all orders of the perturbation theory are considered. Since this is not generally possible, in order to make a reasonable choice for μ_R and μ_F , a scale characteristic for the process under investigation is usually considered [36].

²⁰ $1 \text{ GeV}^2 \leq \mu_F^2 < Q^2$.

Chapter 2

Searching for Dark Matter

As illustrated in the previous chapter, it is undeniable that the SM of particle physics is one of the most successful theories in physics and the recent precision tests and the discovery of the Higgs boson have firmly confirmed its validity [3]. However, despite this success, the large number of free parameters (25 or 26)¹ that need to be manually defined suggests that the SM is not the ultimate theory. Indeed, many unanswered questions remain within the SM, hinting at the existence of yet undiscovered physics commonly referred to as *physics beyond the SM* (BSM).

For example, an open question is the so-called *big hierarchy problem*: the SM does not account for the difference between the electroweak scale ($O(100 \text{ GeV})$), the scale at which electroweak and strong force become equally strong (10^{16} GeV), and the *Planck scale* (10^{19} GeV) at which also the gravitational interaction becomes as strong as the other forces.

Another big challenge is the *matter anti-matter asymmetry*; at the Big Bang stage, matter and anti-matter are supposed to have been created in the same amount and it is expected to still find this proportion today. In reality, matter is predominantly observed in the universe and there is a huge shortage of anti-matter. This suggests that an asymmetric annihilation process has taken place. These observations could be explained by assuming a level of CP violation² not compatible with the current formulation of the SM.

Related to the CP violation, there is another unexplained aspect in the Standard Model:

¹These parameters are: the masses of the twelve fermions, the three coupling constants that describe the strength of the gauge interactions, the two parameters describing the Higgs potential, the eight mixing angles of the PMNS and CKM matrices. If the parameter representing the CP violation of the strong force is considered different from 0, the number of SM free parameters becomes 26 [3].

²The CP-symmetry ensures that physics laws remain unchanged upon the concurrent exchange of particles with antiparticles (C) and inversion of the spatial coordinates (P) of the system (Sec. 1.1.2).

the strong CP problem. CP violation is permitted also in the strong force but experimental results, i.e. measurements on the neutron elastic dipole momentum, lead to the exclusion of this possibility [40].

Finally, a combination of cosmological and astrophysical observations lead to establish that the SM only accounts for 5% of the universe content in terms of mass and energy. The remainder is attributed to non-luminous *dark matter* (DM), 27%, and to *dark energy*, 68%. A further major criticality of the current formulation of the Standard Model is the lack of appropriate DM candidates.

In this second chapter, the attention will be focused precisely on this last SM criticality as a background for part of the studies presented in this thesis work. Sec. 2.1 presents some of the main evidences that indirectly infer the presence of DM. In Sec. 2.2, a brief review of the possible particles that have been postulated as DM candidates over the years is given. Finally, in Sec. 2.3 the different DM detection techniques are presented.

2.1 Evidences for Dark Matter: more than a century of research

2.1.1 The early stages of Dark Matter studies

In 1687, Isaac Newton published the laws of motion and Universal gravitation. These laws began to be applied to celestial bodies as well, leading the scientific community to suspect the existence of *dark bodies*, or at least *dark areas*, already at the end of the 19th century. Father A. Sacchi's memoirs speak of "dark nebulae": dark masses scattered in a space characterized by gaseous nature [41]. Moreover, thanks to the invention of astronomical photography, the presence of "dark vacant areas" in the sky was inferred and the astronomers began to investigate whether these areas were due to a paucity of stars or to the presence of absorbing matter. The specific term "dark matter" (*matière obscure* in the French original formulation), was coined by the scientist H. Poincaré in 1906 and in that period several studies were undertaken in this field [42]. Until the early 1930s, however, all investigations on the Milky Way and on the stars in the Solar neighborhood had led to the erroneous conclusion that the maximum allowed amount of DM was about half of the total local density. At that time, the astronomers were also convinced that this obscure kind of matter was likely to consist of faint stars (cool and cold stars and solid bodies) [42].

The pioneering works mentioned so far have cleared the way toward modern determinations of the local DM density. In the early 1930s, some astronomers began to focus on galaxy clusters. Among these scientists it is worth mentioning the Swiss-American astronomer F. Zwicky: in 1933 he studied the red shifts of various galaxy clusters and noticed a large gravitational anomaly within the Coma Cluster [43]. Indeed, applying the *virial theorem*

to the Coma Cluster, he managed to estimate its mass and derived the corresponding mass-to-light ratio, combining mass and total luminosity measurements. He found that the mass-to-light ratio for the Coma Cluster was two orders of magnitude larger than the one in the Solar neighborhood. This generated, for the first time, the suspicion that dark matter was present in much greater amount than luminous matter [42].

2.1.2 Rotation curves

During the following years, the *rotation curves* of galaxies, measured by combining optical surface photometry with observations of the 21 cm hydrogen line, began to play a particularly important role in the discovery of DM. The rotation curve represents the circular velocity profile of the stars and gases in a galaxy as a function of their distance from the galactic center. By applying the Newtonian gravity to the stars in the spiral galaxies, it is possible to find the rotational velocity $v(r)$ at a distance r from the galactic center (Eq. 2.1):

$$v(r) = \sqrt{\left(\frac{G M(r)}{r} \right)} \quad (2.1)$$

where G is the Newton's constant and $M(r)$ represents the amount of mass contained inside the orbit. For a fixed amount of mass inside larger orbits, the velocity should decrease like $1/\sqrt{r}$. Instead, studies on several galaxies have shown that $v(r)$ is approximately constant and this means that $M(r)$ is increasing with distance [6].

During the 70s, K. Ford and V. Rubin, performed investigations concerning the rotational curve of the Andromeda Galaxy (M31) and further studies were carried out also by K. Freeman [44], D. Rogstad and G. Shostak [45]. They considered several different galaxies, such as M33, NGC 2403, IC 342, M101 and NGC 6946. All these researches came to the conclusion that the investigated rotational curves decline slowly at large radii; thus, these galactic objects have flat rotation curves out to radii larger than their optical extent, as shown in Fig. 2.1. These observations show that the galactic masses continue to grow beyond the region occupied by the stars and gases, suggesting that the galaxies might be surrounded by a much bigger DM halo [42].

The rotation curves are extremely compelling and provide one of the most convincing evidences of the DM existence.

2.1.3 Gravitational lensing and bullet cluster

Another indication of the presence of DM comes from *gravitational lensing*. From General Relativity it is known that the gravitational potential produced by massive objects creates a

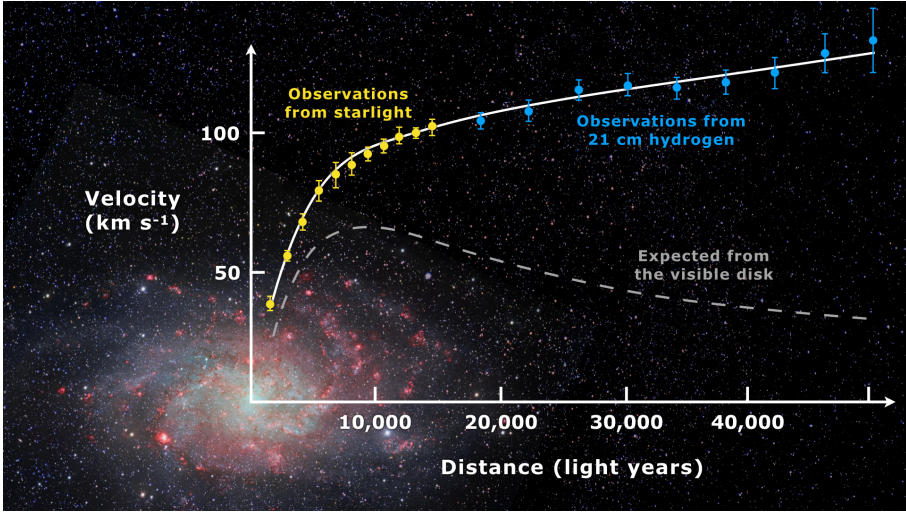


Figure 2.1: Rotational curve of the M33 Galaxy. The top curve shows the observed velocities while the bottom curve the velocities that would be expected considering only the mass of the luminous disk [46].

well in the space that can even bend light passing through and the more massive the object is, the stronger is the deflection. This effect is sketched in Fig. 2.2: light from a galaxy behind a huge mass is bent around forming rings and arc structures when it is observed from the Earth. A concrete example of this effect is given in Fig. 2.3a, where a big cluster galaxy is represented and blue arcs are visible.

In this context, an indirect evidence of the existence of DM might be found by looking at the distances and the masses involved. According to the known theory of gravitational lensing, it is in fact discovered that there is not enough visible matter to justify the observed effect. Fig. 2.3b shows a blue halo superimposed to the visible matter: this represents the amount and distribution of postulated DM that would be needed to explain the measured lensing effect. Those portions of the picture where there is no correspondence between visible matter and simulated DM could be considered the proof for the existence of DM.

Another display of the presence of DM comes from the “*Bullet cluster*” formed by the collision of two galaxy clusters (Fig. 2.4).

The normal baryonic matter (represented in pink) is detected as X-ray emission, while the blue regions show where the highest mass concentration is found as inferred via gravitational lensing. The baryonic matter is not primarily centered around the mass distribution so the blue regions can be interpreted as the dark distributions. In Fig. 2.4 it is possible to notice that the dark matter exhibits a spherical distribution, whereas the baryonic matter presents a distortion. This behavior can be explained by considering that the self-interacting hot gas

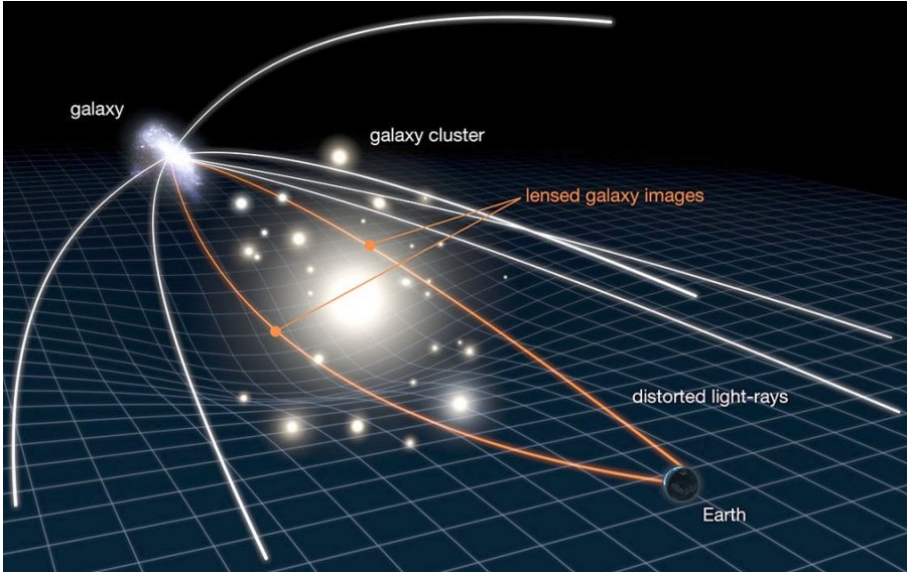
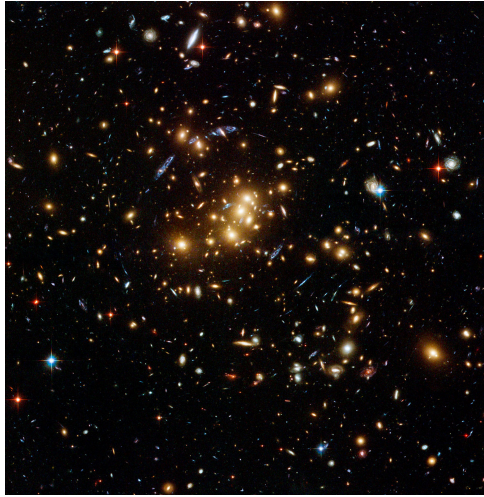


Figure 2.2: The light from distant galaxies behind a large galaxy cluster can be bent and distorted by its gravity. The representation provided in the picture is deliberately out of scale to easily visualize the effect [47].



(a)



(b)

Figure 2.3: (a) The galaxy cluster ZwCl 0024+17; (b) Ring of dark matter around ZwCl 0024+1652 [48, 49].



Figure 2.4: The *Bullet Cluster* is produced by the collision of two galaxy clusters. Pink areas correspond to the location of hot gas, detected as X-rays while blue regions correspond to the *dark matter halo* inferred from gravitational lensing [50].

forming most of the clusters' baryonic mass decelerates; on the contrary, the DM shows a spherical distribution suggesting that it did not undergo any interaction with the gas nor with itself and was therefore not slowed down.

2.1.4 Simulations

More recently, the advent of supercomputers and the significant increase in computing power over the last decades, allowed to simulate the evolution of the matter distribution in the universe. By comparing simulated data to large observations surveys, it is possible to understand the physical processes underlying the buildup of real galaxies. These simulations have shown that it is not possible to model the structures observed in the universe if only visible matter is included. On the contrary, a closer agreement between simulated data and observations is obtained if DM is considered [51, 52].

2.1.5 Cosmological Microwave Background

The evidence for the existence of DM on scales of the size of galaxies clusters does not provide an estimation of the total amount of DM in the universe. Cosmological and astrophysical measurements related to the large scale structures of the universe make it possible to perform such estimations.

The universe emerged about 13.7 billion years ago in an event called *Big Bang* [53]. Atoms were formed about 380000 years after the Big Bang, during the *recombination* phase, when the universe was cool enough³ for protons and electrons to bind without constantly being ionized by photons [54]: at that early stage, the universe became transparent for photons. These photons are still measurable today as a *microwave signal* that comes from every direction and forms the so-called *Cosmic Microwave Background* (CMB). This radiation was predicted in 1948 by G. Gamow and his collaborators [55] and it was discovered by R. Wilson and A. Penzias in 1965 [56]. Today it is known that the CMB is characterized by a black body spectrum with a temperature of 2.7255(6) K [57]. The CMB anisotropy, manifesting itself as small temperature fluctuations of the order of $\delta T/T \simeq 10^{-5}$, provides very rich information on the early universe, allowing the calculation of cosmological parameters, such as the abundance of matter [56]. The fluctuations can be represented with gaussian peaks (Fig. 2.5) referred to as *acoustic peaks*.

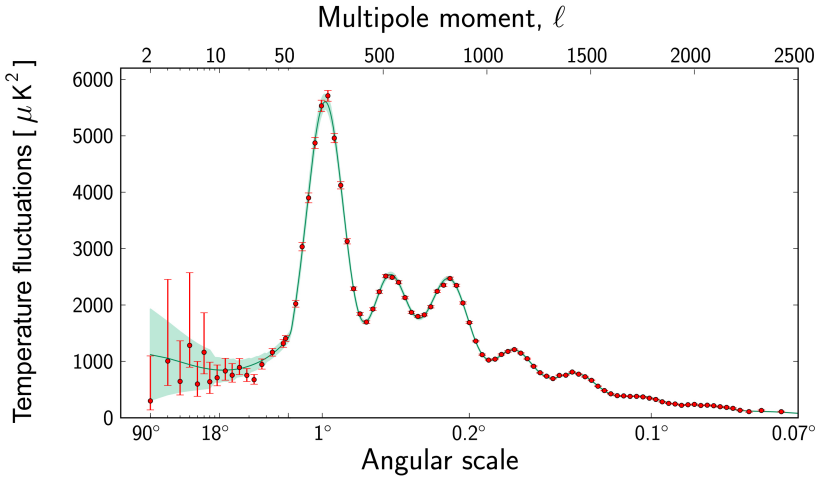


Figure 2.5: CMB temperature fluctuations as measured by the PLANK satellite [58, 59].

The position, shape and relative height of the different peaks provide important informa-

³ $T = 3000 \text{ K} \approx 1 \text{ eV}$.

tion:

- the shape of the first peak, placed at $\ell \sim 220$ and angular scale $\sim 0.9^\circ$, accounts for the density of standard baryonic ($\Omega_b h^2$) and cold dark matter ($\Omega_c h^2$) which are two of the primary fit parameters⁴. Their values, extrapolated from the fit are $\Omega_b h^2 = 0.02207(33)$ and $\Omega_c h^2 = 0.1196(31)$, respectively⁵;
- the second peak is suppressed compared to the other two peaks; this can be explained introducing a certain amount of dark baryons [58];
- the existence of the third peak combined with its relative height provides further information on the dark matter density [56].

As mentioned, considering the best fit values of the spectrum, it is possible to estimate the composition of the matter-energy content of the universe: 4.9% is baryonic ordinary matter, the dark matter contributes for 26.8% and 63.3% is dark energy. This means that there is ~ 5.5 times as much dark matter as ordinary matter in the universe.

2.2 Dark Matter candidates

By the early 1980s it was clear that this shortage of dark matter could not be due to ordinary matter in the form of white dwarfs, neutron stars or cold clouds of gas but had to be a new form of matter [6].

So far, none of the observations (or simulations) involving DM is able to give a clear indication of what DM is made of. However, investigations on the large structures of the universe together with previous collider results have provided some constraints on the nature of dark matter that will be summarized below:

- DM particles do not interact electromagnetically, otherwise they could be observed with standard astronomical procedures; also, their sensitivity to the strong force is very unlikely. In general, it is possible to consider DM models which are exceptions to this rule. However, since DM particles susceptible to strong force have not yet been experimentally observed and since these interactions would make these particles easier to detect, this kind of models are disfavored. On the contrary, a wide range of theories postulate the possibility for DM candidates to interact with the ordinary

⁴Densities are expressed in terms of h^2 , a quantity related to the Hubble parameter H_0 , where $h = H_0/(100 \text{ km}/(\text{s Mpc}))$. $H_0 = v/d$ indicates the expansion rate of the Universe (where v is an object's velocity and d is its distance from Earth).

⁵At 68% Confidence Level.

matter via the weak force. For this reason, the DM candidates hypothesized by this type of models are called *Weakly Interactive Massive Particles* (WIMPs), as will be discussed later on (Sec. 2.2.3).

- The viable DM candidates have to have masses and cross-sections suitable for yielding the observed relic density, i.e. the abundance of DM present in the universe today equals $\Omega_c h^2 \simeq 0.12$ [60].
- The DM particles must be *stable* on cosmological time scales since, otherwise, they would have already decayed.
- By means of simulations, it has been possible to establish that, in order to reproduce the astrophysical observations, i.e. the large-scale structures of the universe, it is necessary to assume that most of the dark matter was *non-relativistic* (cold) at the time of structures formation. However, part of the DM could also be warm (with velocities not quite relativistic) and hot (with relativistic velocities) [61].
- DM, in general, should be *collisionless*. An evidence for this aspect can be found in the collision of two galaxy clusters, i.e. the *Bullet cluster*. As explained, the dark matter component of the two clusters passes through each other unhindered.

2.2.1 Neutrinos

Considering the constraints on the nature of DM particles, there is only one potential DM candidate among the fundamental particles of the SM: *the neutrino*. This possibility was first investigated in 1976 by A. S. Szalay and G. Marx [42]. Neutrinos are stable and, unlike all the other SM particles, do not experience electromagnetic and strong interactions: these are essential features for almost any viable DM candidate. However, they are essentially ruled out as the dominant (or only) dark matter components for two main reasons: 1) their abundance is not large enough to account for the amount of DM derived from observations; 2) neutrinos are also too relativistic (too hot) to result in the structures formation observed in the universe [62].

Although SM neutrinos are far too light and hot to make up the DM, this did not preclude the possibility that other types of neutrino-like particles could constitute this elusive matter. In 1993, S. Dodelson and L. Widrow proposed *sterile neutrinos* as possible DM particles: an additional neutrino species that could have been produced in the early stages of the universe and that does not experience the electroweak interactions⁶ [63]. Other than through gravity, this new particles are allowed to mix with the SM neutrinos and, due to this mixing, their sterility is not perfect so that they can decay into SM photons and neutrinos [61, 64]. Investigations on the possible decays have provided stringent constraints

⁶This requirement can be achieved by assuming that sterile neutrinos are right-handed.

on the sterile neutrinos masses ($\sim 1 \text{ keV} - \sim 10 \text{ MeV}$) and couplings; according to [65], these constraints still allow to consider the sterile particles as potential candidate for all observed non-baryon dark matter. However, sterile neutrinos are extremely challenging to detect given the extremely faint interactions with the other SM particles [66].

2.2.2 Axions

As mentioned, strong evidence is available that the SM of particle physics is not complete. Many theories beyond the SM require new elementary particles, some of which also have the properties needed to be eligible DM candidates. *Axions*, another category of postulated DM particles, are an example. QCD allows CP violation in the strong interactions similarly to what happens for the weak interactions. However, experiments show that the strong force perfectly obeys CP symmetry: for example, one of the CP violation effects would result in the electrical dipole momentum of the neutron being $\sim 10^{10}$ times larger than what experimental upper bounds allow.

In the 70s, R. Peccei and H. Quinn [67, 68] proposed a promising solution introducing a new $U(1)$ symmetry that is spontaneously broken and, later on, F. Wilczek and S. Weinberg [69, 70] argued that a broken symmetry implies the existence of a new boson, which they called *axion*.

A combination of laboratory and astrophysical investigations placed strong constraints on the axions mass ($m_{\text{ax}} \lesssim 10^{-3} \text{ eV}$) and on their interactions with SM particles, which are much weaker than originally envisioned by Wilczek and Weinberg [71, 72]. The assumptions on the axions' production mechanisms introduce large uncertainties on the relic density estimation; however, there still exist possibilities for these candidates to comply with all current constraints [57].

2.2.3 Weakly Interacting Massive Particles

Although there is a broad consensus that the dark matter in the form of SM neutrinos would not be able to account for the observed universe's large scale structures, these particles provide an important template for the class of postulated species known as *Weakly Interacting Massive Particles* (WIMPs). This term, coined by G. Steigman and M. Turner, was initially intended to include all DM candidates (including axions, gravitinos, etc.), whereas it currently only denotes those particles hypothetically interacting via the weak force.

WIMPs, often indicated as χ , are characterized by a mass of the order of a few GeV to TeV and by cross-sections comparable to those of the electroweak interactions. They are stable and neutral [64]. The DM relic density observed today can be explained by reasonably

assuming that these particles were produced in thermal and chemical equilibrium with the hot soup of SM particles after the inflation [57]. This appealing property, referred to as the WIMPs miracle, made them the most popular candidates to become DM particles. For this reason, WIMPs have motivated an extensive experimental program that continues to this days (see Sec. 2.3).

Several categories of BSM theories that also include WIMPs can be defined and it is beyond the scope of this thesis to review them all. The main characteristics of supersymmetric theories will be summarized below, and the attention will then be focused on more generic and alternative models postulating a *mediator* particle capable of interacting with both SM and DM particles.

Supersymmetric theories, SUSY, were proposed as an extension to the SM during the '70s [73]. According to these theories, each fermion has a boson *superpartner* and vice versa. These superpartners share with their SM counterparts the same quantum numbers with the exception of the spin, which differs by $\frac{1}{2}$. In some simpler supersymmetric theories, the supersymmetric sector also shares masses with the corresponding SM particles.

The semantic operation that is applied to indicate this new sector of particles is the following: the suffix “-on” of the SM bosons is replaced with the new suffix “-ino” (e.g., the SM photon becomes the DM *photino* and the gluon becomes a *gluino*). This procedure is also applied to W , Z and the Higgs boson which then they become *Wino*, *Zino* and *Higgsino*, respectively. The bosonic DM partner of the SM fermionic component is instead indicated by prefixing the name with the letter “-s”, e.g. the SM electron becomes the DM “*selectron*”, neutrino becomes DM “*sneutrino*”, etc.

Supersymmetric theories are related to the WIMP-paradigm since most of them assume the existence of WIMP-like particles. For instance, the lightest stable SUSY particle (LSP) matches accurately the WIMPs features. In the simplest supersymmetric extension of the SM, called *Minimal SuperSymmetric Standard Model* (MSSM) it is assumed that the *R-parity*, defined as $P_R = (-1)^R$, is conserved. R is a quantum number and it is equal to $3(B-L)+2s$ where B , L and s represent baryonic, leptonic numbers and spin, respectively. It follows that SM particles have $P_R = 1$, while for the supersymmetric candidates $P_R = -1$. The requirement of the R-parity conservation results in the impossibility for the LSPs to decay, thus making them excellent candidates for DM particles. In the MSSM theory there are two possible candidates as LSP: the sneutrino and the neutralino, but the former has already been ruled out by more recent experimental results. Other supersymmetric theories have other LSP candidates matching the characteristics of the WIMPs; this is the case, for example, of the *gravitino* (the superpartner of the graviton) and the *axino*, the axion's superpartner.

Alternative models to SUSY are more generic theories that are interpreted in the framework of *simplified DM models* [74, 75] and that postulate the possibility of interactions between

Standard and Dark Matter through *mediator* particles [76]. These models, as presented in Sec. 2.3.3, are especially relevant for collider experiments.

The simplest models of this category of theories hypothesize that these interactions between visible matter and DM occur through existing particles, such as Z weak interaction boson (*Z portal model*) or Higgs boson (*Higgs portal model*). Indeed, it is assumed that these SM bosons, in addition to the ordinary decay models involving SM particles, could decay into two dark particles.

Additionally, there are theoretical models that conceive new types of dark boson mediators that are not included in the SM and that could couple to a dark sector that may include one or more dark candidates, such as WIMPs. In these models, the dark matter mediators are unstable heavy particles produced directly from the interactions of SM particles, such as quarks, and they are able to decay into both SM and DM particles. These simplified dark matter models can be classified according to the properties of the particles that mediate the DM-SM interaction, i.e. the *mediator sector*.

There is a first class of such models where the mediator sector is composed of an extended Higgs sector and an additional mediator. On the contrary, another group of models postulates the existence of a single particle that acts as an interaction mediator [76]. The nature of this mediator varies from model to model; indeed, it can be both scalar with spin-0, pseudo-scalar with spin-0, or vector-boson or axial-boson (both with spin-1), or it could also be a spin-2 particle [77]. This last class of simplified models, which assumes the existence of a single mediator, will be further explored in Chapter 7 since the model used as a benchmark in the Trigger Level Analysis (TLA) belongs to this category.

In general, the simplified models are not structured to the point of being considered actual theories, but this is not necessarily a negative aspect of this approach. Indeed, the true nature of dark matter is currently unknown and all the models of the DM interactions are mostly speculative within their constraints. Moreover, if the history of the DM discovery will follow the same path of the SM, it is likely that simpler processes will be discovered first [51].

2.2.4 MACHOS

MACHOs, acronym for Massive Astrophysical Compact Halo Objects, denote other class of DM candidates in response to the leading alternative of WIMPs. MACHOs are non luminous or faint astronomical objects mostly localized in the galaxy halo and characterized by a too low brightness to be detected by current telescopes. This category includes objects such as primordial black holes and brown dwarves. Today, there is a large consensus that MACHOs cannot account for the main component of DM [6].

2.2.5 MOND

All the argumentations presented so far are based on Einsteinian or Newtonian gravity. The so-called Modified Newtonian Dynamics (MOND), proposed by M. Milgrom in 1982 [78, 79], aims to extend the Newton's laws of gravity instead of postulating the existence of the dark matter. At the heart of the MOND theory is the recognition that if instead of obeying Newton's second law $F = ma$, the force due to gravity was scaled⁷ as $F = m \frac{a^2}{a_0}$, it would be possible to account for the observed motions of stars and gases within galaxies without postulating the presence of DM. MOND allows to explain many observations on galactic scale (e.g. galactic rotation curves) without introducing DM [57]. However, this theory is not compatible with the general theory of relativity. Attempts to embed MOND into a relativistic framework require the existence of additional fields (e.g. a vector field or a second metric) and introduce considerable arbitrariness.

2.3 WIMPs searches

Theoretical models of DM provide valuable information on the possible ways in which DM particles interact with ordinary matter. These suggestions from theory turn out to be crucial in the experimental field: indeed, they are fundamental for the design and construction of detectors sensitive to DM and allow to develop appropriate analysis methods if any product of DM interactions were recorded.

Three general categories of experiments that search for WIMPs signals can be identified:

- *direct search*, based on the detection of nuclear recoils in a target volume (see Sec. 2.3.1);
- *indirect search*, aiming to detect the annihilation products of the WIMP pairs (see Sec. 2.3.2);
- *collider search*, where the presence of DM would be detectable as a signal of missing transverse energy⁸ (Sec. 2.3.3), or as resonances on the invariant mass spectrum (Chapter 7).

These categories are complementary to each other and all rely on the existence of interactions between DM and SM in addition to gravitation.

⁷This correction is valid in the limit of very low accelerations ($a \ll a_0 \sim 10^{-10} \text{ m/s}^2$).

⁸Missing transverse energy (E_T^{miss}) quantifies the energy that is not detected but is expected due to energy and momentum conservation laws. In a p-p collision, the total energy involved in a hard scattering process is not known. However, the fraction of momentum perpendicular to the beam axis is expected to be equal to 0. In case non-detectable particles such as neutrinos (but also more exotic particles like WIMPs) exist, the E_T^{miss} value obtained considering detectable particles will be different from 0.

2.3.1 Direct search

Underlying all WIMPs direct detection experiments is the assumption that WIMPs are gravitationally trapped inside galaxies with an adequate density profile capable of explaining the observed rotation curves (Sec. 2.1.2). This type of experiments assumes that WIMPs, characterized by masses ranging from 10 GeV to 10 TeV, interact through elastic scattering on nuclei by depositing energies ranging from 1 to 100 keV (Fig. 2.6). The *expected inter-*

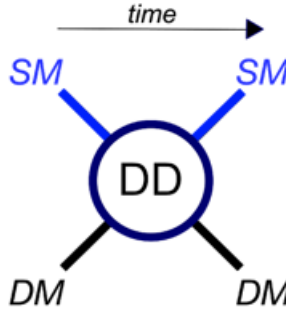


Figure 2.6: The diagram shows $DM \rightarrow SM$ particle scattering targeted by *direct detection* experiments [51].

action rates depend on the product of the local WIMPs *flux* and the *interaction cross section*. The WIMP flux can be estimated considering the local density of DM ($0.39 \text{ GeV}/\text{cm}^3$), the mean WIMP velocity, similar to that of the stars that populate the galaxy and estimated around 220 km/s, the galactic escape velocity (544 km/s) and the WIMP mass, which is currently not fixed. Like the mass, also the WIMP interaction cross-section is an unknown parameter, so far: this explains why the limits on the WIMP experimental observables are presented as a function of this degenerate mass-cross-section pair.

Knowing that the center of the galaxy is stationary with respect to the distribution of WIMPs and that the Sun and Earth move with respect to the center of the galaxy, the DM halo of the galaxy is not rotating; thus the Earth’s motion through this halo produces an apparent “wind” of WIMPs. This wind has two distinct signatures in the WIMP-induced recoil signal:

- strong daily forward/backward asymmetry of the nuclear recoil direction; this effect is due to the changing direction of the Earth passing through the WIMP cloud.
- annual modulation of the recoil rate due to the motion of the Sun in the galaxy and the motion of the Earth. There are times of the year when the velocities of the Sun and the Earth are in agreement, others when they are discordant: the highest

agreement is expected around June 2nd and the lowest around December 2nd. The effect of this seasonal modulation is small (5%): the Earth's orbit is at a large angle with respect to the galactic plane (about 60°) and this attenuates the modulation effect.

WIMPs induce low energy (~ 10 keV) nuclear recoils and these processes can release energy by *ionization*, *scintillation* and *phonon vibrations*. However, the detection of these signals is extremely challenging; indeed, the expected signal rates are far below the typical radioactive background, which requires the laboratories to be placed deep underground. The detectors need to be shielded against residual radiation due to muons or radioactivity from the rocks and to use materials with very low intrinsic radioactivity. Moreover, background charged particles and gammas induce electron recoils, which also deposit energy via the same three channels mentioned above but with different energy-partition. A detector capable of discriminating between electron and nuclear recoils can achieve substantial background reduction since fast neutrons are the only background left.

Present generation WIMPs detectors are mostly hybrid and sensitive to two energy deposition channels at the same time; this features allow to exploit the fact that nuclear and electronic recoils are characterized by different energy deposits, so these devices can effectively discriminate the two types of recoil.

Many experiments aiming to the direct detection of DM have been built: DAMA/LIBRA, CRESST, CoGeNT, CDMS-II, XENON100, LUX and SuperCDMS. A detailed analysis of these experiments is beyond the scope of this thesis. The most relevant results of some of the experiments mentioned will be reported below.

DAMA is the only collaboration that, so far, has claimed a signal exceeding the 5σ level: they observed a modulation of the WIMPs signal with the expected period (1 year) and phase (maximum recorded around 2 June). This result is confirmed by several years of data taking. Also the CoGeNT experiment observed an annual modulation characterized by not negligible significance (2.8σ) and, in general, other experiments such as CDMS-II and CRESST-II have reported positive results observing unexpected number of events in the signal region, although these still need to be confirmed⁹. More experiments, such as XENON, LUX and PICASSO have also produced important data that allowed to set more strict limits on WIMPs-nucleon scattering cross-sections [15].

⁹According to [80] and [81], the results from DAMA/LIBRA and CoGeNT, respectively, should not be considered fully confirmed and/or conclusive.

2.3.2 Indirect search

Indirect DM detection refers to the search for the *annihilation* or *decay fragments* from DM particles (Fig. 2.7).

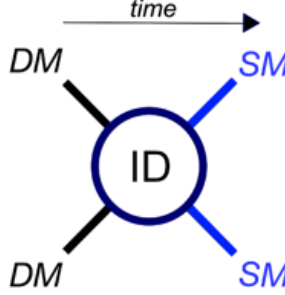


Figure 2.7: The diagram shows DM annihilation to SM particles as sought by *indirect detection* experiments [51].

The results of these processes include detectable species such as gamma rays, neutrinos and anti-particles (i.e. positron, anti-protons and anti-nuclei). The production rate of the final state particle f per unit volume from DM annihilation Γ_f^A can be expressed by Eq. 2.2:

$$\Gamma_f^A = c \frac{\rho_{\text{DM}}^2}{m_{\text{DM}}^2} \langle \sigma v \rangle N_f^A \quad (2.2)$$

where $\langle \sigma v \rangle$ represents the thermally-averaged cross-section for DM annihilation multiplied by relative velocity [82] calculated at the appropriate temperature, ρ_{DM} is the density of DM and N_f^A represents the number of final state particles f produced in one single annihilation event, finally c is a constant value.

As mentioned, DM annihilation to practically any final state produces gamma rays. Fermi Large Area Telescope (LAT) and also ground-based facilities such as HESS, VERITAS, MAGIC and HAWC have provided a detailed picture of how the gamma ray sky radiation should appear to be compatible with DM indirect production. Excess of gamma rays from the direction of the inner galaxy, where a signal from DM annihilation might be especially bright [83], has been claimed by LAT. However, the nature of this excess is quite controversial: while the shape of the spectrum is in agreement with what expected for a standard WIMP with mass of a few tens of GeV [84], unresolved point sources have been supported as possible counterparts [85].

WIMPs can be captured and trapped in celestial objects where their probability of annihilation and their density can be enhanced. This can lead to muon neutrinos production which can interact in the Earth: neutrino telescope such as MACRO, BAKSAN, SuperKamiokande, Baikal, AMANDA, ANTARES, NESTOR and the large sensitive area IceCube can detect the upward going muons.

Stable charged particles produced directly from DM annihilation or decay are abundant in the cosmic radiation and are a prime target for indirect DM searches. To maximize signal to noise ratio, searches focus on rare particle species: for instance positrons, anti-protons and anti-nuclei.

An excess of high-energy positrons has been firmly established by several experiments, in particular by AMS-O2 [86]. The excess has been ascribed to DM annihilation, although other explanations for the excess, albeit controversial and strongly constrained by data, also exist [87, 88]. AMS-O2 has also reported an excess in the anti-proton flux between 10 and 20 GeV and energies above 100 GeV. This could be interpreted as a possible signal of DM annihilation [89].

Anti-nuclei such as anti-deuterium and anti-helium could also be produced by DM annihilation or decay. Specific detector designs have been developed to identify low-energy cosmic-ray anti-nuclei. The detection of even a single anti-nucleus would have considerable importance as a possible sign of new physics [90].

2.3.3 Colliders search

Assuming that WIMPs can annihilate into visible matter or can scatter off nuclei as presented in Sec. 2.3.1 and Sec. 2.3.2, it is also possible to hypothesize that DM particles can be produced by colliding SM particles (Fig. 2.8).

Various searches for DM have been performed using colliders such as the Large Hadron Collider (LHC) at CERN and these type of investigations have targeted different DM models [15, 76, 91]:

- UV-complete models, such as supersymmetry;
- models with rich dark sector;
- models with long-lived particles;
- Effective Field Theories (EFTs) with heavy mediators;
- models that interact with the SM via mediators that can be Higgs or Z boson or new types of dark boson mediators that are not included in the SM.

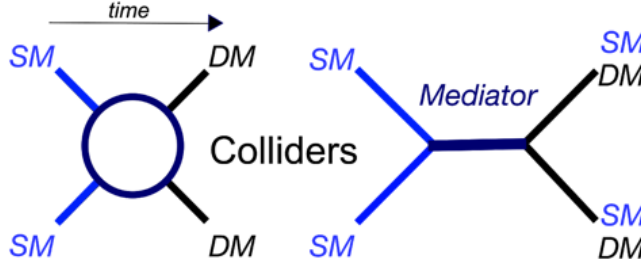


Figure 2.8: Production of DM particles at colliders. The diagram on the left shows the production of DM particles from the annihilation of SM particles. The diagram on the right represents the case in which the interaction occurs through a *mediator* particle between DM and SM particles [51].

The experimental program combines searches for invisible-particle production mediated by SM bosons with more generic searches that involve invisible particles produced via new mediators and finally specific searches based on complete models [15].

DM produced at colliders would almost always pass invisibly through the detector, since the WIMPs are assumed to be stable. However, a variety of signals could highlight the presence of dark matter at colliders [77, 92, 93].

An imbalance in the momentum measured in the transverse plan in an event might indicate the presence of undetectable dark matter particles that have been produced together with one Standard Model particle. To identify DM events, additional particles, such as $X = \text{jet}, \gamma, Z, \text{Higgs}, (t)\bar{t}, (b)\bar{b}$ should be produced in conjunction with DM in p-p collision so that it becomes possible to detect the recoiling WIMPs as missing transverse momentum with magnitude E_T^{miss} ¹⁰. In mono-jet, mono- γ and mono- Z searches, the SM particle is emitted as initial state radiation (ISR); on the contrary, in the other searches different kinds of models are considered (see, for example, [92] for a more detailed overview).

If DM candidates can be produced via an s-channel exchange of a new particle, then this mediator could also decay back into SM final states: resonance searches can therefore also be used to constrain DM models. A *resonance* indicates the presence of an extremely short-lived particle which manifests itself as a peak in the mass distributions, since the cross section increases at a mass value corresponding to that of the particle produced. This leads to an excess of events (a *bump*) in the otherwise smoothly falling di-jet or di-lepton *invariant mass distributions*. This kind of searches will be further discussed in Chapters 7 and 8.

¹⁰Not all invisible decay products are necessarily DM: neutrinos are not detected by detectors at colliders, but the neutrino background can be independently calculated from the SM.

Beyond Standard Model phenomena exist that do not give rise to resonances; for this type of signals it is required to look for an excess of events in the *angular distribution* (instead of the mass spectrum) of a dark matter mediator. The angular distribution of energetic jets relative to the beam axis is represented by the *polar scattering angle* in the two partons CM frame θ^* . QCD calculations predict angular distributions peaked at $|\cos \theta^*|$ close to 1. On the contrary, models of BSM processes predict more isotropic angular distributions. Discrepancies from the expected QCD behaviors could represent an evidence for new physics [94].

In general, the interplay of mass resonance, angular analysis and $X + E_T^{\text{miss}}$ searches depends on the specific theoretical model under investigation [95].

So far, at particle colliders, none of the approaches presented succeeded in observing significant discrepancies with respect to the Standard Model predictions; limits on the dark matter couplings to ordinary matter or limits on the dark matter particles and mediators masses have been set [51].

Chapter 3

LHC and the ATLAS experiment

This chapter provides a description of the LHC accelerator as well as the ATLAS experiment. The Large Hadron Collider's most relevant features are outlined in Sec. 3.1 [96]; in Sec. 3.2 the attention is instead focused on the ATLAS detector. ATLAS (*A Toroidal LHC Apparatus*) is the largest of the four major LHC experiments and is the result of the work of a large collaboration of several thousand physicists, engineers, technicians and students over a period of thirty years [97].

3.1 The Large Hadron Collider

LHC is a two-ring superconducting hadron collider operating at CERN (Conseil Européen pour la Recherche Nucléaire) near Geneva, Switzerland. It is located in the 27 km-long tunnel formerly used for the Large Electron Positron collider (LEP)¹ and 100 m below ground across the Swiss/French border. The design of the LHC was driven by two main objectives: measure the SM parameters in previously inaccessible phase space regions and discover BSM phenomena. The LHC physics program is mostly based on proton-proton collisions which occur at maximum energies of 14 TeV and luminosity greater than $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Furthermore, LHC can also generate collisions of heavy ions (Pb or Xe) at lower energies and luminosity (up to 2.8 TeV per nucleon and $10^{27} \text{ cm}^{-2}\text{s}^{-1}$, respectively).

The two counter-rotating beams (beam 1 clockwise, beam 2 anti-clockwise) only collide in dedicated collision points. Four of these intersections along the ring host the larger LHC experiments²: two general purpose experiments, i.e. CMS (Compact Muon Solenoid) and ATLAS (A Toroidal LHC Apparatus) (see Sec. 3.2), and other two experiments, i.e.

¹LEP has been in operation from 1989 to 2000.

²There are also other smaller experiments that will not be considered here.

LHCb and ALICE (A Large Ion Colliding Experiment), characterized by a more specialized physics program. ALICE is mainly focused on heavy-nucleus collisions, whereas LHCb addresses physics involving bottom quarks. LHCb has been designed to operate at a luminosity of $10^{32} \text{ cm}^{-2}\text{s}^{-1}$, thus requiring a luminosity leveling in case LHC is operating at higher levels. On the contrary, ATLAS and CMS have been designed to operate at the highest luminosities achievable by LHC and to take advantage of the full potential of the collider.

3.1.1 Acceleration mechanism of protons

Before being fed into LHC, protons need to be accelerated, and this acceleration is organized in several stages: in fact, LHC is the last collider in a larger acceleration complex (see Fig. 3.1).

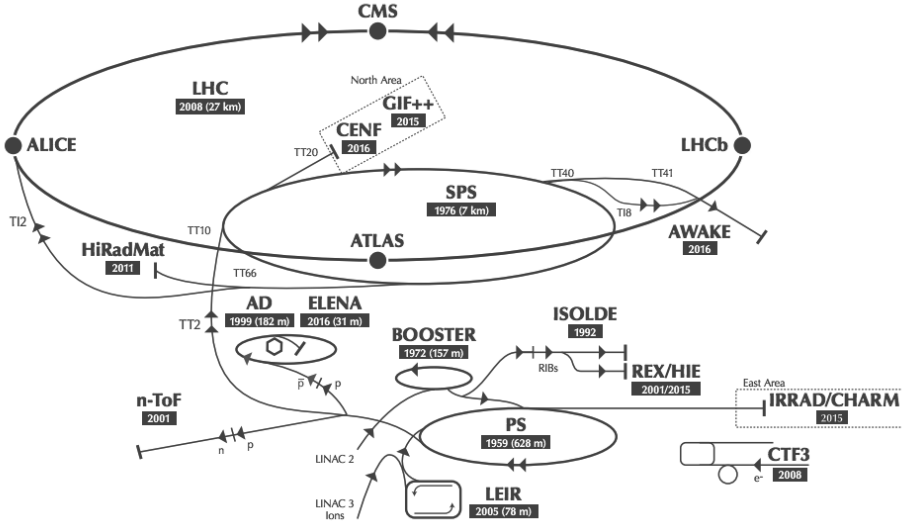


Figure 3.1: Accelerator complex [98].

As a first step insertion chain, the molecules of Hydrogen are stripped of their electrons and the resulting protons are first fed into a linear accelerator, Linac 2, where they are accelerated up to 50 MeV through a series of radio-frequency (RF) cavities. The alternating fields of RF cavities ensure the correct velocity of traveling particles: particles traveling faster (and therefore anticipating the RF oscillations) are decelerated, whereas slower particles are accelerated; ideally a proton with exactly the right energy will neither be accelerated or slowed down.

After Linac 2, the bunches of protons pass through the Booster, where they reach an energy

of 1.4 GeV. The Proton Synchrotron (PS) receives the protons from the Booster and accelerates them to 26 GeV before injecting them into the Super Proton Synchrotron (SPS). In SPS the energies are increased up to 450 GeV, enough to feed the protons into LHC, where they undergo further acceleration by means of eight RF cavities, each of which capable of increasing the beam energy by 16 MeV per round trip. This means that more than 10 million round trips are needed to reach the final energy.

During acceleration, the curvature of the protons' trajectories is matched to that of the beam pipe by means of 1232 superconducting Nb-Ti dipole magnets; 392 quadrupole magnets are in addition responsible of maintaining the beam focus.

3.1.2 Luminosity and pile-up

For a collider, there are two key parameters: the collision energy and the luminosity L , which is a measure of the number of collisions. LHC has been designed to provide unprecedented centre-of-mass energies and luminosity.

The relationship between the event rate R of a given process characterized by a cross section σ is given in Eq. 3.1.

$$R = L \cdot \sigma \quad (3.1)$$

The luminosity only depends on the beam parameters and is defined as (Eq. 3.2):

$$L = \frac{N_1 N_2 b^2 n_b f_{rev} \gamma_r F}{4\pi \epsilon_n \beta^*} \quad (3.2)$$

where N_1 and N_2 represent the number of protons in bunch 1 and bunch 2, respectively; n_b is the number of bunches per beam, f_{rev} represents the reduction frequency, γ_r is the relativistic gamma factor, ϵ_n the normalized transverse emittance, β^* the beta-function at the collision point. The beams collide at a small angle in order to avoid parasitic collisions due to the short distance between bunches; the information related to the crossing angle between beams is encapsulated in the geometric factor F [99]. From Eq. 3.1, it follows that by integrating L over a time interval Δt , it is possible to calculate the integrated luminosity³ \mathcal{L} (Eq. 3.3):

$$\mathcal{L} = \int L \, dt. \quad (3.3)$$

\mathcal{L} accounts for the amount of data N produced in a certain period of time (Eq. 3.4):

³The integrated luminosity is measured using inverse cross section units (fb⁻¹ or pb⁻¹).

$$N = \sigma \int_t^{t+\Delta t} L dt \quad (3.4)$$

The integrated luminosity can be measured experimentally by considering the occurrence of an event with a known cross-section. On the contrary, knowing precisely the value of \mathcal{L} and given the theoretical cross section of rare BSM event, it is possible to estimate the number of time this kind of events might occur. Or again, \mathcal{L} turns out to be crucial to interpret the experimental event count: indeed, if \mathcal{L} is known, information on the cross section can be deduced from the number of events observed.

The increase in luminosity has some disadvantages: in order to maximize luminosity the bunches are squeezed and several proton interactions can occur in each bunch crossing. This phenomenon is referred to as *pile-up*. There are two different forms of pile-up: the *in time pile-up* and the *out of time pile-up*. The first occurs when there are multiple simultaneous p-p interactions in a single bunch crossing, whereas the second arises when signals from previous or future bunch crossing overlap in the detector, which is characterized by a limited read-out time. The pile-up is quantified considering the average number of interactions per bunch $\langle\mu\rangle$; as can be seen from Fig. 3.2, the increase in luminosity determines an increase in the pile-up level and this leads to a deterioration in the resolution of the detector.

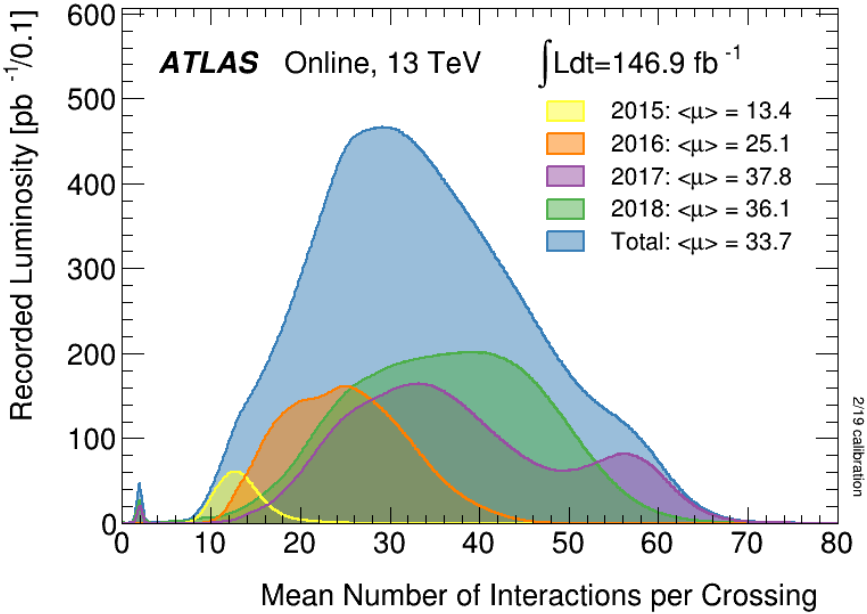


Figure 3.2: The Run 2 pileup distribution [99].

3.1.3 Run schedule

In Tab. 3.1 a summary of main accelerator parameters for the LHC is presented. As shown, the achievement of the design specifications takes place following a gradual, still ongoing, process: periods of data taking (*runs*) alternate with periods of *shutdown* in which both the accelerator and the experiments undergo maintenance and receive upgrades.

Table 3.1: Summary of LHC accelerator parameters [99].

Parameter	Design	Run1	Run2	Run3	HL-LHC
Energy [TeV]	14	7/8	13	14	14
Bunch spacing [ns]	25	50	25	25	25
Bunch intensity [10^{11} ppb]	1.15	1.6	1.2	up to 1.84	2.2
Number of bunches	2800	1400	2500	2800	2800
Emittance [μm]	3.5	2.2	2.2	2.5	2.5
β^* [cm]	55	80	30 \rightarrow 25	30 \rightarrow 25	down to 15
Crossing angle [μrad]	285	—	300 \rightarrow 260	300 \rightarrow 260	TBD
Peak luminosity [$10^{34} \text{ cm}^{-2}\text{s}^{-1}$]	1.0	0.8	2.0	2.0	5.0
Peak pileup	25	45	60	55	150

As can be seen from the data summarized in Tab. 3.1, during the first Run, LHC operated at energies $\sqrt{s} = 7$ and 8 TeV with a spacing between the bunches of 50 ns. After Run 1, at the beginning of 2013, LHC entered the first shutdown phase (the long shutdown 1 or LS1) which lasted about 2 years, during which a great consolidation and maintenance work was done (see for example [100] for more details). The 2015-2018 period was covered by Run 2, where LHC operated at $\sqrt{s} = 13$ TeV and with a bunch spacing of 25 ns.

The second shutdown phase began at the end of 2018. The main purpose of the Long Shutdown 2 (LS2) was the upgrade of the LHC injectors; in addition, full maintenance of all the equipment, consolidation of part of the machine and preparation for the LHC High Luminosity (HL-LHC) project (see for example [101] for more details) have also been carried out.

The HL-LHC operations will begin in 2026 and are planned to continue for the following ten years. In this phase the pile-up in ATLAS and CMS will increase significantly, reaching a value of 150 interactions per bunch crossing. To obtain the design specifications summarized in Tab. 3.1, several updates are required, such as upgrading the injectors to provide higher intensity beam, replacing the focusing magnets in order to further squeeze the beam, and upgrading the various components to be able to support the increased radiation and stored energy [102, 103].

3.2 The ATLAS detector

The ATLAS detector has a cylindrical shape with a diameter of 25 m, a length of 44 m and a weight of 7000 tonnes (Fig. 3.3). As presented in Fig. 3.4, it uses a right-handed coordinate system with the origin placed at the centre of the detector, at the nominal interaction point (IP), with the z -axis directed along the beam pipe⁴. The positive direction of the x axis points from the IP to the centre to the LHC ring, while the y -axis points upwards. The transverse plane is described using cylindrical coordinates (r, ϕ) , where ϕ is the azimuthal angle around the z -axis and r is the radius. The variable η represents the *pseudorapidity*, a convenient quantity expressing the coverage with respect to the beam axis. Pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$, while the angular distance in the pseudorapidity azimuthal angle space is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

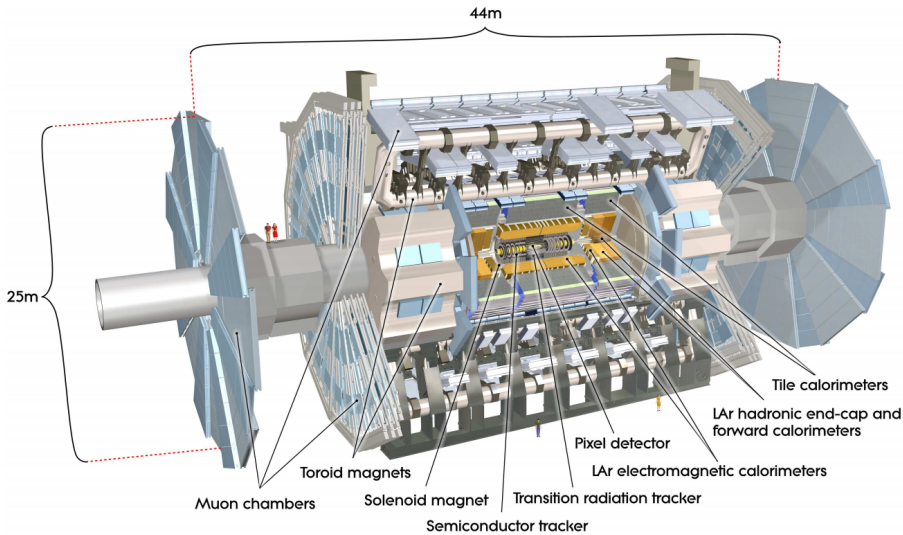


Figure 3.3: The ATLAS detector. Individual subdetectors are identified [97].

3.2.1 The philosophy of the ATLAS design

“To not let anything escape”. With these few words it is possible to summarize the philosophy at the basis of the design and construction of the ATLAS experiment. The detector was conceived to reconstruct and identify all products emerging from the collisions at the LHC and is also an experiment built for the discovery of a wide range of generally new phenomena that were expected to be observed at the TeV scale [97].

⁴The beam pipe axis is different from the beam axis; indeed the beams collide at a slight angle to avoid interaction upstream.

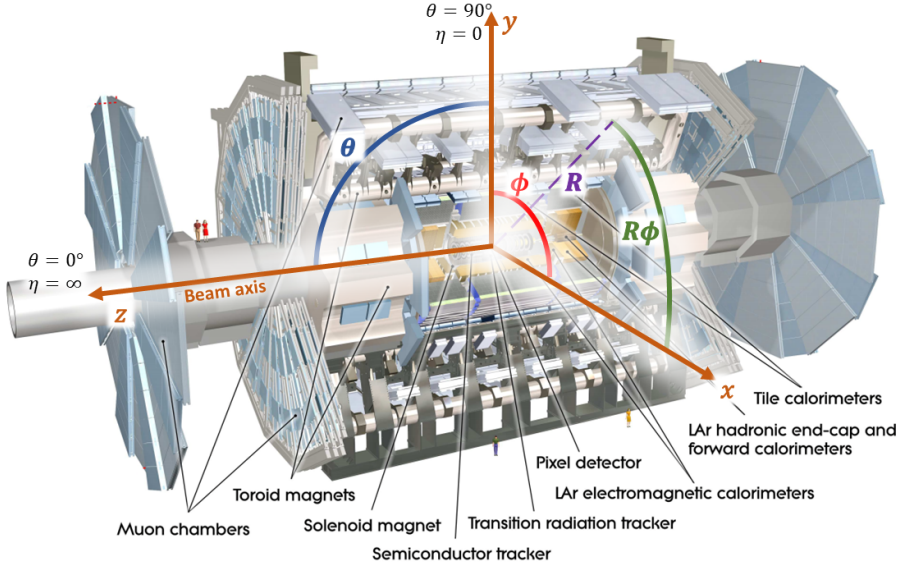


Figure 3.4: Coordinate system of the ATLAS detector. The z axis is directed along the beam pipe, the x axis points towards the center of LHC and the y axis points upwards [97, 104].

The ATLAS detector comprises several sub-detectors (see Figs. 3.3) extending radially around the IP. This *onion-like* design was chosen to allow combining the signals from each individual component to perform particle identification as well as to measure specific quantities, such as energy and momentum, that form the basis of many ongoing studies.

ATLAS is divided into three main sectors: the *tracking detectors* closest to the beam pipe (see Sec. 3.2.3.1 and 3.2.3.2), the electromagnetic (EM) and the hadronic *calorimeters* surrounding the tracking detectors (see Sec. 3.2.3.3) and, on the outermost layer, the *muon system* (see Sec. 3.2.3.4).

In order to select the potentially interesting events, ATLAS uses a *trigger system* (see Sec. 3.2.4) characterized by two different levels; they use increasingly higher granularity information to reduce the data rate down to a few hundred Hz (based on the design luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$).

The detector parts with a cylindrical shape in the central region constitute the *barrel*, whereas the regions placed at the two ends of the barrel are called *end-caps*. In Fig. 3.5, a cross-sectional view of the ATLAS detector is shown; a representation of how the different particles interact with the sub-detectors, thus allowing their identification, is also provided.

For instance, charged particles are the only particles capable of leaving a trace within the inner detector parts: photons and electrons originate EM showers in the EM calorimeter, while the heavier particles such as neutrons and protons mostly undergo hadronic interactions. Muons barely interact with any of the material and are therefore able to reach the outermost part of the detector; neutrinos do not interact with any of the materials and escape undetected. As discussed in Chapter 2, their presence can be inferred from missing transverse energy. This is just an oversimplified example of the fundamental principles of particle interactions in the ATLAS detector; a more detailed discussion on this topic is provided in [105, 106].

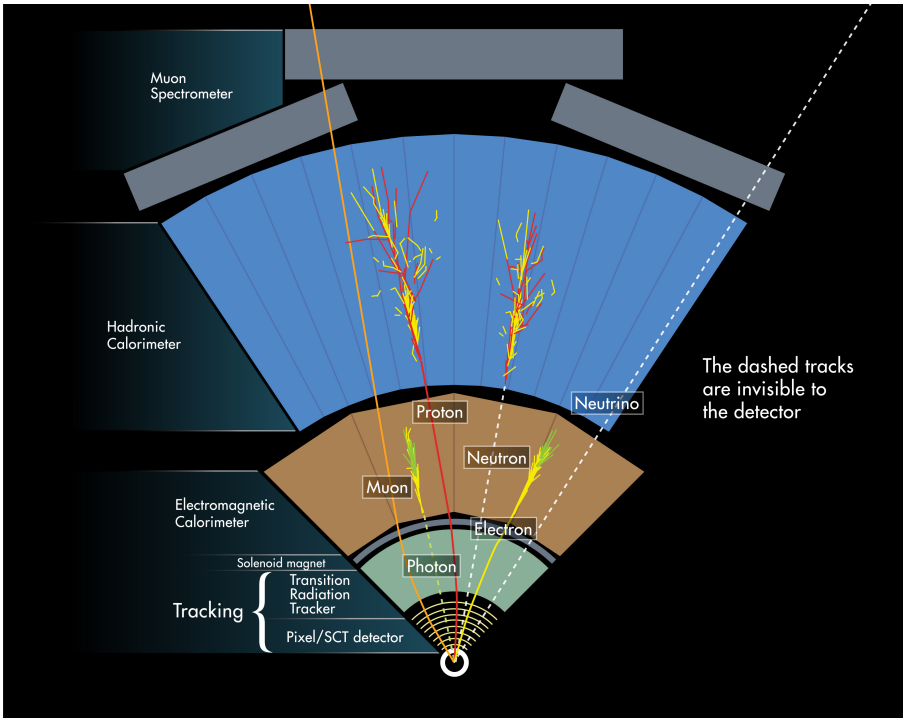


Figure 3.5: Principles of particle identification: a schematic representation of the interactions occurring within a portion of the ATLAS detector is given [107].

3.2.2 Requirements

ATLAS is intended to serve as a general purpose detector, capable of performing a broad set of experiments: from precise measurements of the SM parameters to the search for new physics phenomena. To achieve this ambitious goal, the most important requisites it must fulfill are:

- good charged-particle momentum resolution and reconstruction efficiency in the inner tracker;
- large acceptance in pseudorapidity with an almost full azimuthal angle coverage;
- very good EM calorimetry for electron and photon identification and measurements. These features need to be complemented by full-coverage hadronic calorimetry for accurate jet and missing transverse energy measurements;
- good muon identification and momentum resolution over a wide range of momenta;
- highly efficient triggering on low transverse momentum objects with sufficient background rejection;
- due to the experimental conditions at the LHC, the detectors require fast, radiation-hard electronics and sensor elements. In addition, high detector granularity is needed to handle the particle fluxes and to reduce the influence of overlapping events [97].

3.2.3 The ATLAS detector sub-systems

This section provides an in-depth review of the ATLAS sub-detectors, with particular focus on the Transition Radiation Tracker (TRT) detector and on the calorimeters, being the most relevant for the results presented in this thesis work.

3.2.3.1 Tracking system

ATLAS' innermost detectors (ID) perform the reconstruction of charged particles trajectories, i.e. the tracks, and identify the position of the interactions, i.e. the vertices. The tracking system originally consisted of three different sub-detectors based on different technologies: the pixel detector, the semi-conductor tracker (SCT) and the transition radiation tracker (TRT) (see Fig. 3.6). The measurements of the particles' momenta is based on the curvature of the reconstructed tracks, which is induced by the 2 T solenoid magnetic field surrounding the ID tracking system.

During the Run 1 to Run 2 shutdown, a beampipe and ID upgrade was carried out. In addition to a smaller beampipe, the new 12-million-pixel Insertable B-Layer (IBL) detector was introduced as the first and innermost layer of the pixel detector to further improve the vertex resolution of the existing pixel layers [108].

The *pixel detector* comprises three layers (disks) in the barrel (end-cap) region. Positioned at a radial distance of 5 cm from the z -axis, the innermost pixel layer is the detector part closest to the interaction point. The outermost layer is located at $R = 12$ cm. The pixel detector

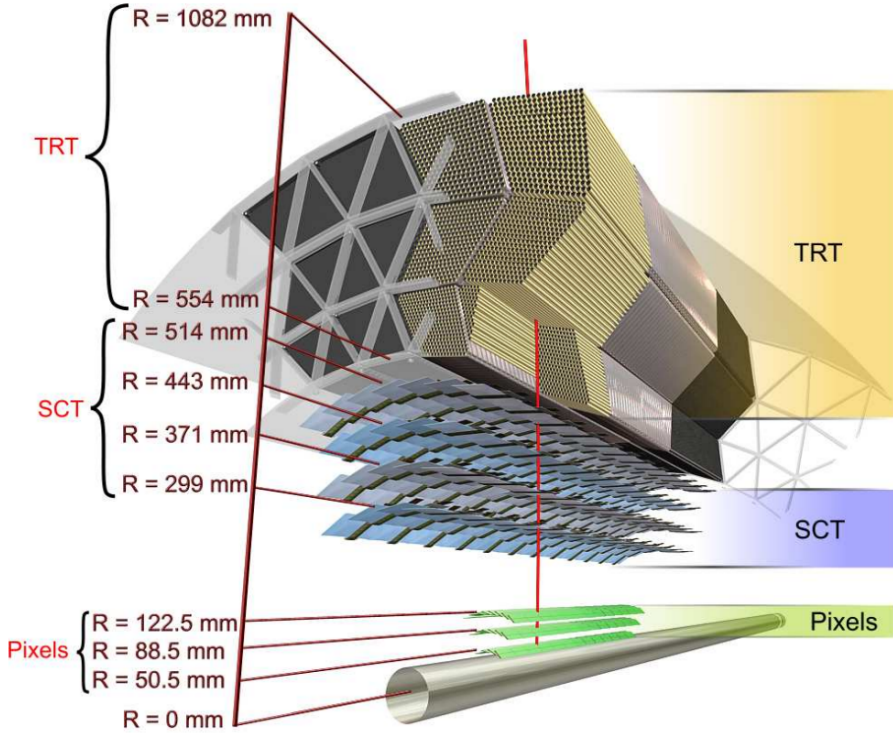


Figure 3.6: Schematic representation of the barrel part of the Inner Detector [97].

provides the highest granularity of the three ID systems. In total, there are 1744 sensors with 46080 read-out pixels each, yielding approximately 80 million read-out channels. The pixels have a minimum size of $50 \times 400 \mu\text{m}^2$ in $(R - \phi) \times z$. In the barrel, an intrinsic accuracy of $10 \mu\text{m} \times 115 \mu\text{m}$ is achieved, while in the end-cap disks tracks can be reconstructed with an accuracy of $10 \mu\text{m} \times 115 \mu\text{m}$ in $(r - \phi) \times R$.

The high precision of the pixel detector is vital for the measurement of impact parameters and for the reconstruction of primary interactions and secondary decay vertices. The latter are needed in order to tag heavy-flavour quarks and τ -leptons via their decays. Here, especially the innermost layer plays an important role. The position of the pixel layers relative to the beam pipe and the other inner detector systems in the barrel region is shown in Fig. 3.6.

The semiconductor tracker (SCT) is a silicon micro-strip, multi-layer detector. Each layer comprises two sets of strips angled with respect to each other to allow 2D position measurement. The barrel region contains four of these layers, with one set of strips parallel to the beam axis. The distance from the z -axis of the innermost and outermost layers is 30 cm and 51.4 cm, respectively. The spatial resolution achieved in the barrel region is $17 \mu\text{m}$ in $(R - \phi)$ and $580 \mu\text{m}$ in z . The end-cap region contains 9 disks with one set of strips

running radially. The spatial resolution is $17\ \mu\text{m}$ in $(R - \phi)$ and $580\ \mu\text{m}$ in R . The 15912 sensors, each containing 768 strips, provide approximately six million read-out channels for the SCT.

3.2.3.2 Transition Radiation Tracker (TRT)

The Transition Radiation Tracker (TRT) is a gaseous ionization detector specialized for electron discrimination. The TRT is composed of three modules: a central barrel and two end-caps. The barrel surrounds the inner silicon tracker and extends radially around the beampipe ($55\text{ cm} \leq R \leq 108\text{ cm}$). The fundamental components of the TRT barrel are the gas-filled straws, parallel to the beam axis, and the radiator material in which they are embedded. The TRT contains a total of 52544 straws, subdivided in 96 trapezoidal modules. The modules are built in three different sizes (referred to as Type 1, Type 2 and Type 3 modules) and are sequentially mounted in $32\ \phi$ sectors. The complete TRT barrel assembly is shown in Fig. 3.7 and a detail of the radial arrangement of the three module types is represented in Fig. 3.8.

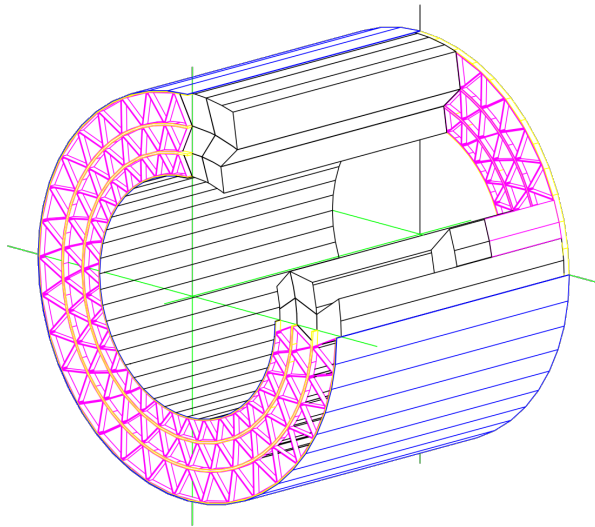


Figure 3.7: The TRT Barrel assembly, showing the Barrel Support System, in pink [109].

A summary of the physical properties of the three module types is presented in Tab. 3.2 [109].

Each module is encased in a carbon-fiber composite cover (*shell*) and is filled with a low-density polypropylene-based material, serving as the transition radiator material. A matrix of holes is stamped into this material to allow the insertion of the straw tubes, which are therefore completely surrounded by the radiator. An exploded view of one of the TRT modules is presented in Fig. 3.9.

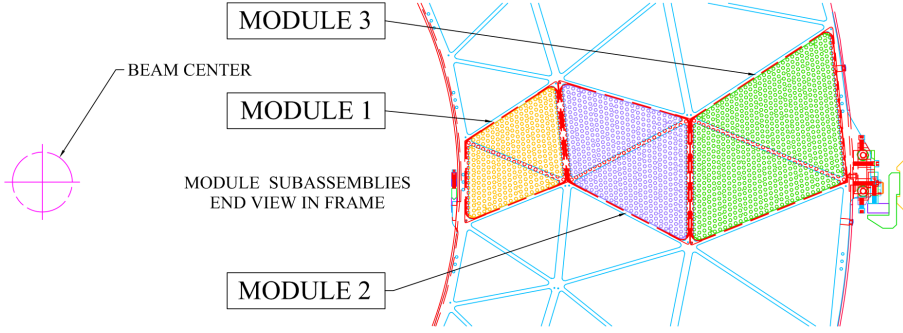


Figure 3.8: Layout of the three types of modules as mounted inside the Barrel Support System [109].

Table 3.2: Summary of the TRT Barrel Module parameters [109].

Module	Inner radius (m)	$ \eta $ at Rmin	Layers	Number of straws	Mass (kg)
Type 1	0.56	1.06	19	329	2.97
Type 2	0.70	0.89	24	520	4.21
Type 3	0.86	0.75	30	793	6.53
Total for barrel			73	52544	439

The radiator material consists of a polypropylene-polyethylene fiber mat about 3 mm thick. Fibers are 19 μm in diameter and are formed from polyethylene clad polypropylene material. The resulting material has a density of 0.06 g/cm³ and the 3 mm-thick pads are stacked perpendicularly to the straws to fill the complete module length. Each full-length module contains approximately 500 pads and the hole layout is represented in Fig. 3.10.

Each straw acts as an independent gas ionization chamber filled with either Xe or Ar, mixed with CO₂ for avalanche quenching. The signal is read out from a thin W/Au anode wire centered inside the straw cathode tube.

3.2.3.3 Calorimeter system

Measurement of electron and photon energies in the range above several hundred MeV as well as hadron energies above 1 GeV is commonly achieved by means of *calorimeters* (Fig. 3.11). The fundamental principle of calorimetric methods is the total absorption of the particle energy in a bulk of material, followed by the measurement of the deposited energy. A calorimetric measurement is inherently destructive; in fact, the majority of particles end their journey in calorimeters.

Two main technologies can be identified for the construction of calorimeters:

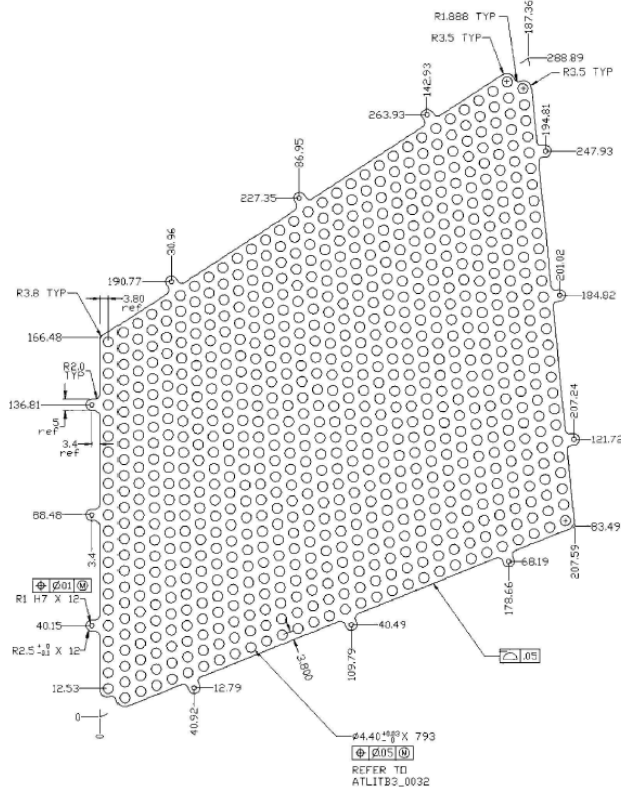


Figure 3.10: A TRT straw alignment plane (divider) showing the layout of the straw holes [109].

Definitions for EM calorimetry The longitudinal and transverse dimensions of EM calorimeters are given by two quantities called radiation length X_0 and Moliere radius R_M , respectively. They describe the evolution of an EM shower and are defined as follows:

- the *radiation length* X_0 represents the mean distance travelled before an electron loses $1/e$ of its energy via bremsstrahlung. An electromagnetic shower can be initiated also by photons; in this case, the mean free path before a photon undergoes pair production is larger than the electron's radiation length: $\lambda_{\text{prod}} = \frac{9}{7}X_0$. It is important to stress that the radiation length is a material-specific quantity and is measured in g/cm^2 .
- the lateral width of an EM cascade is mostly determined by multiple scattering and is best described by the *Moliere radius* $R_M = \frac{21}{E_c} \frac{\text{MeV}}{X_0}$, where $E_c = \frac{500 \text{ MeV}}{Z}$ is

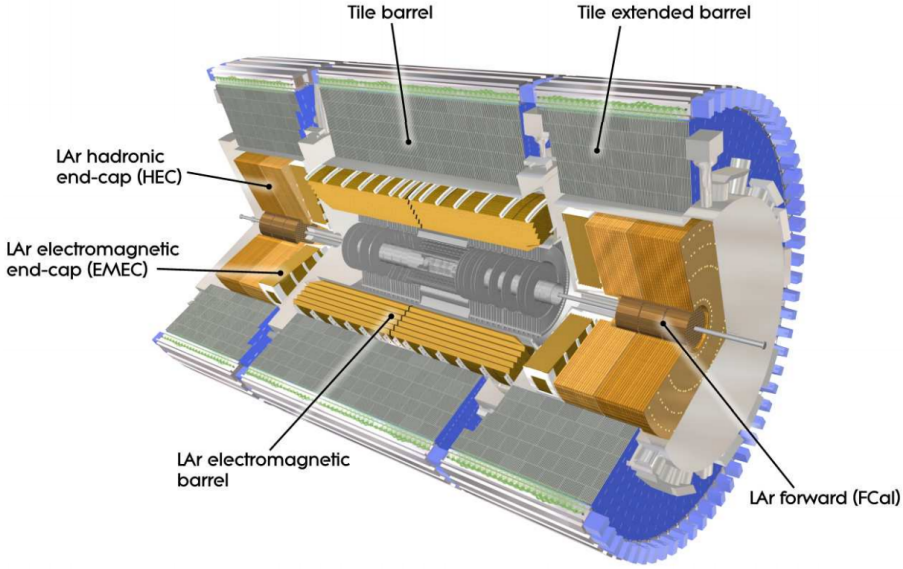


Figure 3.11: Cut-away view of the calorimeter system in ATLAS [97].

referred to as *critical energy* and represents the lower threshold for the multiplication of shower particles to occur. On average, 95% of the shower energy is contained within a cylinder around the shower axis with radius $R(95\%) = 2R_M$, irrespective of the energy of the incident particle.

Barrel LAr geometry As anticipated, the barrel section of ATLAS houses both EM and hadronic calorimeters. The LAr EM calorimeter is built from lead absorber plates and liquid argon active layers and is therefore based on ionization. The calorimeter's elements are laid out in an accordion shape providing full ϕ coverage. The total length of the LAr exceeds $22X_0$ to minimize energy leakage and the entire assembly is placed outside the solenoid. A representation of the LAr geometry is shown in Fig. 3.12.

Definitions for hadronic calorimetry Hadronic calorimeters share the same fundamental working principles of their electromagnetic counterpart. Unlike EM calorimeters, however, the longitudinal shower development is no longer measured in unit of radiation length X_0 , but is instead characterized by the average nuclear interaction length $\lambda_I \approx 35 \text{ g/cm}^2 \text{ A}^{1/3}$. Indeed, the hadronic energy loss is based on inelastic hadronic processes rather than atomic. The transverse development of an hadronic cascade, caused by large transverse momentum transfers, is significantly larger than the EM case.

The most abundant secondary particles detected in an hadronic calorimeter are neutral

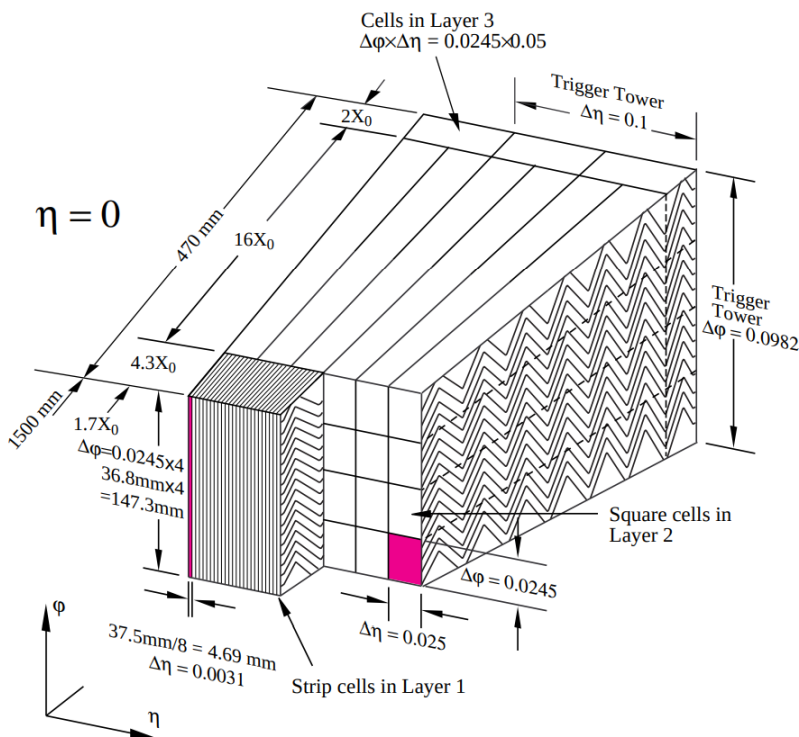


Figure 3.12: View of a LAr calorimeter barrel module. The granularity decreases radially. Layer thicknesses are given in radiation lengths units [97].

and charged pions; kaons, nucleons and other hadrons are produced, albeit with lower multiplicities.

Neutral pions represent approximately one third of the pions produced in each inelastic collision, but their production is subject to large fluctuations determined by the properties of the first inelastic interaction. Once neutral pions are produced, they rapidly decay into energetic photon pairs, thereby initiating EM subcascades within the initial hadronic shower.

Unlike electron and photon energies, which are almost completely recorded in a EM calorimeter, a substantial fraction of the hadronic cascade energy remains undetected. This can be ascribed to the following reasons:

- some part of the hadron energy is required to break up nuclear bonds and does not contribute to the visible energy;
- the extremely short-range nuclear fragments resulting from the break-up of the nu-

clear bonds are generally absorbed before reaching the active layers and, therefore, remain undetected;

- in contrast, stable neutral or extremely long-lived particles have the highest probability to escape the calorimeter, therefore depositing a very small fraction of their energy. This further reduces the amount of visible energy to detect.

Approximately 30% – 40% of a hadronic cascade’s energy remains invisible. It is important to stress that only the EM energy and the energy loss of charged particles (ionization losses) are detectable in a calorimeter. This results in a significant difference in the calorimeter response to electrons and hadrons. A calorimeter is defined as compensating if the relative detection efficiency of hadronic and EM energy deposits, denoted h/e , is unity. Compensation can be achieved by means of low-Z or fissile materials to increase the probability of converting the otherwise lost hadronic energy into visible energy. The ATLAS hadronic calorimetry is non-compensating [105].

Barrel tile geometry The Tile calorimeter is hadronic calorimeter for the ATLAS barrel section. As shown in Fig. 3.13, it is composed of alternating layers of steel absorbers and scintillating polystyrene as active material. Readout is done by means of photomultiplier tubes matched to wavelength-shifting fibers, which collect the UV light emerging from the scintillating layers.

Overall, the barrel and extended parts of the Tile calorimeter allow to achieve $|\eta| < 1.7$ coverage, whereas the average radial depth is $7.4\lambda_I$.

Forward calorimeters Liquid-argon technology is used for both the hadronic (HEC) and EM (FCal and EMEC) calorimeters in the ATLAS forward region. The HEC is built with copper absorber plates instead of lead and has a depth of about $12\lambda_I$; the FCal uses copper for the EM calorimetry in the inner layers and tungsten in the outer layers for the hadronic part. It is positioned at very small angles with respect to the beam axis and has a total depth of about $10\lambda_I$ (hadronic) and $200X_0$ (EM). The EM End Cap (EMEC) adopts the same technique as the EM barrel and covers the $1.375 < |\eta| < 3.2$ region.

Energy resolution of ATLAS calorimeters Electromagnetic calorimeters are mainly characterized by energy and position resolution for photons and electrons. The energy resolution σ_E/E is determined by technical and architectural limitations as well as inherent physical factors, such as the fluctuations of the energy leakage. The characteristics of calorimeters is the improved energy resolution for increasing energies: $\sigma_E/E \propto 1/\sqrt{E}$.

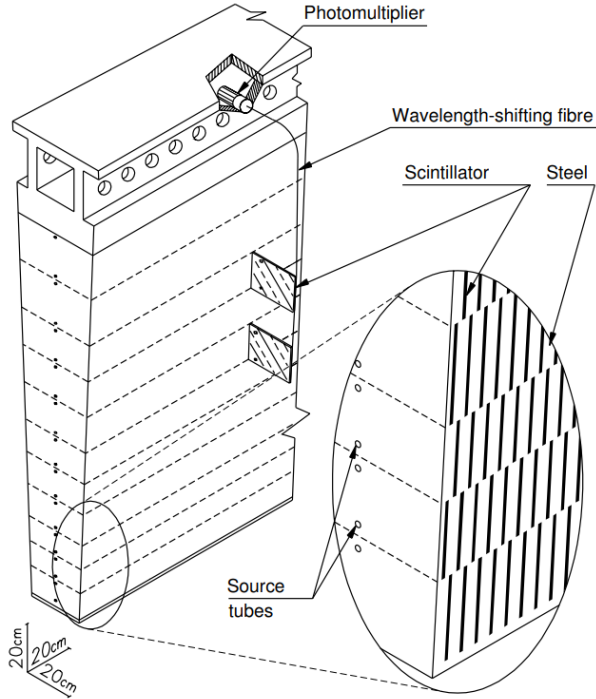


Figure 3.13: Detail of a single hadronic calorimeter tile [97].

As discussed, the fundamental principle of a calorimetric measurement is the conversion of all the incoming particle's energy into detectable form, such as UV or visible light. Therefore, homogeneous calorimeters allow to completely absorb the deposited energy and have the best energy resolution. Sampling calorimeters, on the other hand, are characterized by a more limited energy resolution, due to sampling fluctuations. Although characterized by the same energy dependence, the energy resolution in hadronic calorimeters is reduced even further, due to the limitations in the possibility of detecting hadrons.

A summary of the ATLAS calorimeter's energy resolutions is presented in Tab. 3.3; limitations arising from the operating conditions (noise, pile-up effects, etc.) are not included in the presented table.

3.2.3.4 Muon system

The outermost layer of the ATLAS cylinder is occupied by the muon system, consisting of chambers arranged in three layers. The chambers in the barrel region are cylindrical and parallel to the beam axis, whereas those in the end-cap regions are planar and perpendicular to the beam. Track coordinates are measured by Monitored Drift Tube chambers (MDTs),

Table 3.3: Summary of ATLAS calorimeters energy resolutions.

Calorimeter	Energy resolution $\sigma E/E$			
EMB	10%	$/\sqrt{E}$	\oplus	0.7%
EMEC	10%	$/\sqrt{E}$	\oplus	0.7%
HEC	50%	$/\sqrt{E}$	\oplus	3%
FCal	100%	$/\sqrt{E}$	\oplus	10%
Tile	50%	$/\sqrt{E}$	\oplus	3%

which cover most of the detector regions within $|\eta| < 2.7$. Additional coverage in the forward region $2.0 < |\eta| < 2.7$ is achieved by means of Cathode Strip Chambers (CSCs) located in the innermost layer. CSCs are multi-wire proportional chambers (MWPCs) where the cathodes are divided in orthogonal strips.

The objective of all the sub-detectors within the large muon system is the discrimination of charged particles, which are the signature of the passage of muons. Indeed, at the typical radii of the ATLAS muon chamber, the probability of detecting charged particles not corresponding to muon events is very small.

All the components of the muon system are embedded into the magnetic field provided by the superconducting toroid magnets. Such field, orthogonal to the trajectories of the traversing muons, is crucial for the momentum measurements based on the muon tracks. The magnet system consists of three 8-coil magnets: the large barrel toroid for the $|\eta| < 1.4$ region and the two end-cap toroids for the $1.6 < |\eta| < 2.7$, which form a 22.5° angle with respect to the barrel.

Constant monitoring of the muon system relative alignment is crucial for an accurate momentum reconstruction.

3.2.3.5 Forward detectors

In addition to the components of the main ATLAS cylinder, more sub-detectors are located in the forward region, along the beampipe axis at increasing distances from the interaction point:

- The Luminosity measurement using Cherenkov Integrating Detector (LUCID-2), located at $z = \pm 17$ m from the interaction point, provides a real time measurement of the instantaneous luminosity. The forward inelastic proton-proton scattering is detected by means of small Cherenkov detectors equipped with thin quartz windows as Cherenkov medium.

- The Zero Degree Calorimeter (ZDC), located at $z = \pm 140$ m from the interaction point, consists of two sampling calorimeter modules (tungsten is used as the absorber and quartz rods as the active medium). It detects neutral particles up to $|\eta| \leq 8.2$, contributes to the heavy-ion collision centrality measurement and, during low-luminosity proton-proton collisions, it can provide triggering of minimum-bias events.
- The most distant sub-detector from the interaction point is the Absolute Luminosity For ATLAS (ALFA) detector, located at $z = \pm 240$ m. The detection of scattered protons at very small angles allows to measure the total proton-proton cross section as well as the luminosity; it also provides calibration for the LUCID detector. The ALFA detector consists of four Roman pots with scintillating-fiber trackers and it can be moved as close to the beam as 1 mm.

3.2.4 Trigger system

During LHC Run 2 operations, the event rate is 40 MHz, corresponding to one bunch crossing every 25 ns. Storing the entire amount of data produced by each event (each event is characterized by multiple collisions) is not technically feasible. It is therefore necessary to implement a system, the trigger system, capable of selecting only the potentially relevant events, while promptly discarding the excess of data. Such system can comprise several levels for increasingly specific event selection.

The trigger of the ATLAS detector comprises two levels:

- the Level-1 (L1) trigger selects high-momentum physics objects (e.g. electrons, photons, muons and jets) based on the signals from the EM and hadronic calorimeters. The L1 trigger reduces the event rate from 40 MHz to 100 kHz and, in addition, defines Regions of Interest (RoI) in the $(\eta - \phi)$ -space characterized by large energy deposits (or high-momentum muons) occurred. These regions are fed to the following trigger levels and allow to further reduce the amount of data to handle;
- the High Level Trigger (HLT) is the software-based, second trigger level for ATLAS. It is responsible for reconstructing physics objects from the RoIs selected at the L1 level. The resulting event rate after the HLT selection is 1 kHz, which is $\sim 10^5$ times smaller than the LHC event rate.

Lower energy events occur more frequently than higher energy ones. The trigger system, therefore, selects lower-energy events more often, but these also carry less interesting (and typically already known) physics. To mitigate the effect of the reduced triggering rate for higher energy events, a *prescale factor* N_p is applied. Thus, only one in every N_p events

selected by the trigger are recorded; if $N_p = 1$, the trigger is referred to as *un-prescaled*. The real triggering rate can be recovered by multiplying by the prescale factor. From a statistical standpoint, prescaling represents a limit to the achievable precision for lower energy events.

Trigger Level Analysis (TLA) applied to dijet resonances takes advantage of the L1-fulfilling objects already reconstructed at HLT level and, in order to benefit from the un-prescaled event rate, only stores a limited amount of data per event. A more in-depth description of the trigger system is presented in Chapter 8.

3.2.5 Data Handling

A data taking run in ATLAS, identified by a unique number, can only start if all the detector's components are ready for acquisition. A run is divided in luminosity blocks, which can be flagged according to different quality criteria. The *good runs list* (GRL) contains a summary of the luminosity-related information and is intended for use by physics analysis.

The ATLAS control software Athena is responsible for digitization and reconstruction of both real and simulated data [110]. Only after a successful selection of an event at the trigger level, the full detector is read out. One event in this raw byte-stream format is about 1.6 MB and requires further processing to be used for physics analysis.

The processing of the LHC data is handled by the worldwide network of computing resources known as Worldwide LHC Computing Grid (WLCG), or, simply, *Grid*. The Grid is structured in layers, or *tiers* identified by a level number: the higher the level, the more refined the data. At Tier-0, CERN's lowest in-house level, the first calibrations are applied and Event Summary Data (ESD) is produced by reconstruction algorithms: the data stream is reduced to about 1 MB. Analysis Object Data (AOD), containing physics information needed for the analysis, is obtained from ESD and the size is reduced by a factor of 10 to approximately 100 kB. Each tier acts as a decentralized storage layer for the underlying level as well as a provider of more refined data for universities, laboratories and institutions. The last level, i.e. Tier-3, is the closest to the end users.

The physics analysis in ATLAS is based on the reconstructed and identified objects produced during p-p collisions [111]. For each particle there are specific reconstruction algorithms and it is beyond the scope of this thesis to review them. Since the analysis presented in this thesis is focused on the search for low mass di-jet resonances, a brief overview of the jets reconstruction will be presented in Chapter 7.

3.2.6 Monte Carlo Simulation

Computer simulations are an essential component of the ATLAS experimental activity. In order to accept or reject a theory or to set limits on the probability of observing a given physical process, actual observations must be matched to expected results.

The software framework adopted for the ATLAS experiment is called Athena and is controlled by Python scripts steering C++ algorithms and objects. The Athena framework is based on the C++ Gaudi framework, originally developed for the LHCb experiment, and relies as much as possible on the CLHEP common libraries [112].

The typical workflow of the ATLAS simulations is presented in Fig. 3.14 and it is subdivided in three phases:

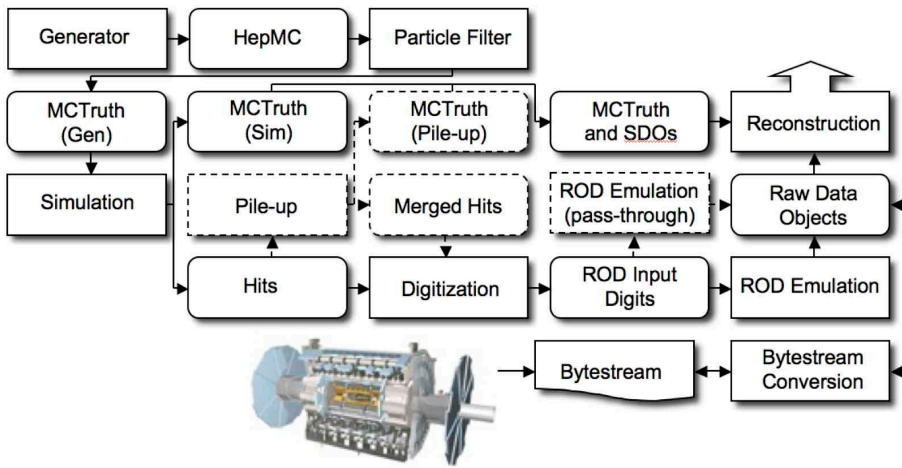


Figure 3.14: Data flow of the ATLAS simulation software, from generators to reconstruction [112].

- **Event generation:** Quantum Chromodynamics (QCD) describes the quark and gluon interactions occurring during a proton-proton collision (see Sec. 1.2). Such interactions are exceedingly complex to be calculated analytically from theories. An event generator provides a list of all the particles emerging from a collision between two protons and, in particular, of those particles with half-lives $> 10^{-11}$ s. Indeed, the most short-lived states would have no probability of reaching any of the detector elements and can therefore be omitted from any subsequent calculation. For this reason, the event generator can be unaware of the detector geometry. There are different general-purpose generators available for LHC physics such as SHERPA [113] and PYTHIA [114, 115]; more specialized tools, such as MadGraph (a matrix element generator) and Herwig++ [116] (specifically used for generating many BSM models

with full spin correlations) are also available. The so-called *truth*, a record of all the particles produced by the generator together with the history of all the interactions, is kept in addition to the output fed to the simulations, which is saved in the standard HepMC format [117].

- **Detector simulation:** once the event generation phase is complete, the particles emerging from the primary collision must be studied individually and their interactions with all the detector components need to be evaluated. Simulations of muon sub-detectors in ATLAS are performed with Geant4 [118–120]⁵. *Fast simulations* can instead be adopted for calorimeters and for the inner detector. All the energy deposits occurring throughout the detector are recorded as *hits*, where a hit contains total energy deposit, position and time. The simulation data flow is conceived to obtain an output from the simulation chain identical to the output collected from the actual measurements. This allows to seamlessly feed observed and simulated data to the same ATLAS trigger and reconstruction packages [117]. As for the event generation phase, the truth is also saved for simulations; in this case, tracks and decays of certain particles are stored. The aim of these truth records is to provide a term of comparison for quantifying the effectiveness of the reconstruction software, as they provide the exact information that should be reconstructed.
- **Digitization:** in general terms, the output of a detector simulation consists of energy deposits, times and locations within the detector. This data must be converted to quantities resembling the actual detector outputs: times, voltages and currents, typically. In addition, all the relevant detector effects must be added to the pure results obtained from the models. Charge distributions in silicon, drift of ionized gases, background from cosmic rays, very low-energy physics processes, etc. are some examples of what is folded in during the digitization phase. Pile-up can be simulated at this stage as a post-processing step to reduce the simulation's required CPU time. The first level trigger is simulated in *pass* mode, meaning that no events are discarded, but each trigger hypothesis is evaluated. The final output has the form of a Raw Data Object (RDO) file, which can be fed to the High Level Trigger (HLT) and to the reconstruction.

The separation between generation, simulation and reconstruction adds flexibility to the data flow: very fast generation jobs can provide several thousands of events, which can be stored and replicated for different versions of the simulation code. As an example, it is possible to postulate the existence of a new particle. From theory, its fundamental properties, such as half-life, decay products, interaction cross-sections with matter and with other particles, etc. can be derived or inferred from models. These properties are then compiled into instructions for the event generators and the physics treatment is implemented in Geant4.

⁵A detailed description of the Geant4 toolkit is given in Chapter 4.

This allows to simulate energy deposits in the detector. Lastly, the digitization step adds all the detector-specific effects, common to all measurements. The set of simulated data is finally compared to observations and the existence of the postulated particle can either be confirmed, disproved or excluded with a certain Confidence Level.

ATLAS simulations are the most time consuming step, so it is usually possible to carry out only simple jobs on standard computers. Distributed computing on the WLCG is the preferred solution to this problem. Software is constantly updated and deployed to all the Grid's nodes, together with a minimal set of data files to allow quick validation jobs.

On the Grid, a simulation with 500000 $t\bar{t}$ events would split in several jobs, each running on a single CPU for the maximum allowed time of 2-3 days. The output is then saved by the ATLAS Distributed Data Management system (DDM) which allows users to search datasets and perform data analysis on them. The Grid software controls also provide management of the job queues and can automatically pause and resume complex jobs depending on the availability of the input and output data.

Part II

Optimizing the ATLAS simulation workflow

Introduction to Part II

The upgrades in view of the HL-LHC runs set ambitious computational requirements: ATLAS expects to record data at a rate of 10 kHz, which is approximately ten times larger than the 1 kHz achieved so far [121]. This increase in the production of measured data requires a matching increase also in the production of simulated data, which is already taking up almost 40% of the CPU hours consumed by the ATLAS experiment. An intensive Research and Development effort is therefore ongoing to upscale the ATLAS simulation resources in order to achieve tangible speed-ups and a large number of different paths are currently being investigated, from parallel computing, to GPU-based tools, to a more intensive use of fast simulations.

In the following chapters, the attention is focused on two aspects of the performance improvement effort: optimizations achievable during the building process of the simulation code and increased efficiency due to the use of newer shape definitions for the detector geometry construction.

Chapter 4 presents an overview of the Geant4 simulation toolkit and provides the fundamental background to compiler optimization and different build types. In Chapter 5, the effects of different optimization options and build types on a standalone Geant4 simulation execution time are discussed. Lastly, in Chapter 6, the studies on the optimization of the ATLAS Transition Radiation Tracker (TRT) geometries are described.

The studies presented in this work provide a valuable baseline for future improvements. The potential of build-time optimizations is quantified and the data collected suggest that the extension of such optimizations to the Athena-based simulation code can introduce significant performance improvements. Also, the effectiveness of optimizing the geometry detector descriptions is confirmed and the observed improvements indicate the opportunity to extend such studies beyond the single TRT.

Chapter 4

Monte Carlo simulations and software optimizations

Particle and detector simulations are a fundamental step in the workflow of the ATLAS detector. In this chapter, the Geant4 simulation toolkit is presented. The concept of a Monte Carlo simulation is outlined, then the main characteristics of Geant4 are discussed. Unlike other simulation codes, Geant4 libraries and user applications need to be built from source code. The fundamental concepts of building and optimizing code are therefore reviewed in preparation for Chapter 5, where they will be applied to a standalone Geant4 benchmark simulation.

4.1 Monte Carlo simulations with Geant4

4.1.1 The Monte Carlo method

The Monte Carlo method is a statistical sampling technique suitable for providing numerical solutions to a broad range of scientific problems. Initially applied to the study of neutron diffusion and multiplication problems in fission devices, the Monte Carlo method has proven to be effective in many areas of physics and mathematics [122].

Monte Carlo simulations rely on the sampling of random numbers in order to make the choices for advancing the calculation. For instance, considering the problem of the random flight of a neutron, it is necessary to reproduce the average behaviour according to which a fraction $\exp(-x/\lambda)$ travels a distance $\geq x/\lambda$ mean free paths without colliding. This cannot be achieved by sampling random numbers from a uniform distribution, but an

exponentially decreasing probability distribution function is instead required.

At the basis of the Monte Carlo method is the requirement of obtaining an arbitrary probability density function $q(x)$ starting from the uniform distribution $u(x)$, for which many efficient random number generators are available. The transformation is represented in Fig. 4.1.

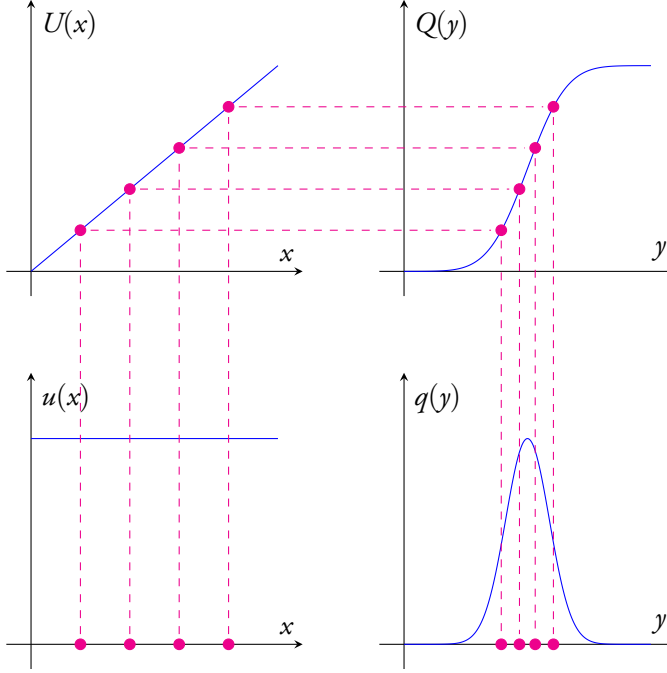


Figure 4.1: Transformation of a uniformly distributed random variable described by the probability density function (PDF) $u(x)$ into a non-uniformly distributed variable described by $q(y)$ [123]. $U(x)$ and $Q(y)$, defined in Eq. 4.3 and Eq. 4.4 respectively, are the Cumulative Distribution Functions (CDFs) for $u(x)$ and $q(y)$. The CDF $U(x)$ is calculated from the PDF $u(x)$, then the inverse of the CDF $Q(y)$ is calculated to sample values distributed according to $q(y)$.

The Cumulative Distribution Function (CDF) of a random variable X represents the probability that a sample drawn from the variable distribution is smaller than or equal to a given value x (Eq. 4.1).

$$\text{CDF}(x) = \text{Prob}(X \leq x). \quad (4.1)$$

Its corresponding Probability Density Function (PDF) is the derivative of the CDF (Eq. 4.2).

$$\text{PDF}(x) = \frac{d}{dx} \text{CDF}(x). \quad (4.2)$$

Let u and q be two monodimensional Probability Distribution Functions (PDFs) describing the random variables X and Y , respectively. Let x be a sample from X . The objective is to transform the sample x in a sample y from Y distributed according to q .

Let U be the Cumulative Distribution Function (CDF) for the random variable X :

$$U(x) = \int_{-\infty}^x u(t) dt \quad (4.3)$$

and, similarly, the CDF for Y is:

$$Q(y) = \int_{-\infty}^y q(t) dt \quad (4.4)$$

If x is a sample from X and y a sample from Y , it is possible to define their statistical equivalence imposing:

$$U(x) = \text{Prob}(X \leq x) = \text{Prob}(Y \leq y) = Q(y) \quad (4.5)$$

In other words, x and y are statistically equivalent if they split the corresponding distributions in the same way [123].

It is now straightforward to conclude that:

$$y = Q^{-1}(U(x)) \quad (4.6)$$

Numerical integration methods and interpolations are required to handle CDFs without an analytical form.

Once a given physical process is known and can either be represented by analytical models or by observed data, PDFs representing the process' behaviour can be obtained. Uniformly generated numbers can then be transformed into statistically equivalent samples drawn from the non-uniform distributions representing the investigated physics.

Therefore, in order to perform a Monte Carlo simulation, the behaviour of the interacting particles and materials must be known in advance, either from measurements or postulated from theory.

4.1.2 Geant4 structure

Geant4 is a Monte Carlo toolkit for particle simulations written in C++ [118–120]. Geant4 provides models for particle transportation through matter and its modular design suits many fields of application, such as high energy, nuclear, accelerator and medical physics.

Unlike other Monte Carlo codes which provide a pre-packaged executable that can only be configured at runtime within a limited range of options, Geant4 consists of a set of libraries which must be assembled by the user into a specialized executable.

The Geant4 codebase is extensive and provides classes to describe the geometry required for the simulation, the physics that should be applied, the properties of the interacting particles, etc. The main drawback of the lack of a default set of options is that the user is required to provide a very detailed description of the problem. On the other hand, the typical modularity of an object-oriented code allows the definition of custom interaction points for the user. Indeed, as it will be discussed in the following sections, handles are provided to retrieve data at increasing level of detail, from coarse summaries to step-by-step information from each interaction of a particle.

A Geant4 simulation is conceptually composed of three main blocks [124]:

- description of the simulated geometry in terms of solids and materials;
- definition of the simulated particles and of the physics that should apply to the problem;
- implementation of data extraction code to retrieve the results of the simulation.

4.1.3 Geometry and material description

A physical entity is fully qualified when *shape*, *properties* and *context* are provided.

In Geant4, *shapes* are defined from the abstract `G4VSolid` class, which is inherited by several classes describing specific solids, such as boxes or cylinders. At this stage, solids are decoupled from their location in space, so the only requirement is to specify the dimensions. More complex solids can also be created directly in the C++ code as boolean solids, i.e. from the union or intersection of simpler objects.

Once all the solids are created, it is necessary to add material and magnetic field specifications and to define whether or not a given solid is a *sensitive detector*. Sensitive detectors in Geant4 are volumes capable of firing a trigger each time they are accessed or traversed by a particle. All these additional properties are matched to a solid and form an instance of the `G4LogicalVolume` class.

The last step is to provide a context to the newly defined logical volume. This is achieved by positioning the logical volume in space and by establishing a mother-child relationship between different volumes. The outermost volume, defining the boundaries of the simulated space, is referred to as *world volume* and is the mother volume to all the other volumes. A single logical volume can be matched to an individual physical volume or to a set of replica or parameterized volumes.

This method of defining geometries provides a consistent interface to the user, but for large geometries can quickly become inefficient, due to the large amount of coding required. To facilitate the ingestion of complex geometries, to allow centralized storage of shared parameters and dimensions and to enable runtime changes, Geant4 offers native support for GDML-based geometries ¹ [125]. These definitions can be created outside Geant4 by means of Computer-Aided Design (CAD) programs, ingested and translated during the detector construction phase and handled seamlessly by the application. This feature will be further discussed in Chapter 5.

As discussed, logical volumes are the objects responsible for matching shapes and materials. Materials can be described in two different ways: an interface to the National Institute of Standards and Technology (NIST) material database is implemented in the `G4NistManager` class and allows to retrieve predefined materials, elements and all the corresponding isotopes [126]; alternatively, the `G4Element`, `G4Isotope` and `G4Material` classes can be used to define custom materials starting from their fundamental constituents.

4.1.4 Physics lists and modules

The physics treatment of a Geant4 simulation comprises:

- the definition of the physical processes for each particle taking part in the simulation;
- the choice of the primary particles to be generated.

The definition of the physics context for the simulation is done by the so-called *physics lists*. A physics list is an object specifying all the particles that are involved in the simulation together with the list of the physical processes that they can undergo.

This approach, specific to Geant4, provides flexibility: users can choose which particles to include or exclude and which physics should be simulated for them. By design, Geant4 does not provide a default description of the physics. It is important to stress that, given the

¹Geometry Description Markup Language, or GDML, is an application-independent geometry description format. It serves as a geometry implementation language and, additionally, provides a geometry data exchange format.

broad range of applications for which Geant4 is suitable, it would not be possible to provide a general and equally accurate physics description for all the particles at all the energies. Moreover, even within the same field of application, different levels of approximation can be justified, e.g. due to performance requirements, and the modularity of Geant4 physics lists facilitates this fine tuning. On the other hand, omission of relevant particles or interactions can lead to wrong simulation results and Geant4 does not provide any safeguard against this risk.

Physics lists in Geant4 can either be constructed from fundamental principles by implementing the abstract `G4VUserPhysicsList` class or by using modular or reference physics lists. Modular physics lists allow to define the physics treatment by choosing modules, each handling a specific category of physics, e.g. electromagnetic, hadronic, decay, etc.; reference physics lists are instead ready-to-use libraries and each is pre-packaged with different combinations of physics processes. When such lists are chosen, it is the user's responsibility to validate them. Detailed documentation about physics lists is available at [127].

4.1.5 Geant4 simulation workflow

In Geant4, a *track*, represented by the object `G4Track`, describes the state of a simulated particle at a given instant in the form of a snapshot, i.e. without keeping any record of the previous particle history.

Tracks are propagated on a step-by-step basis, where a *step* (represented by the `G4Step` object) provides delta information about the change of the state of a simulated particle. A step is composed of a *pre-step* point and a *post-step* point. These store information such as position, direction, energy, material, volume, etc. Steps are concatenated in a chain-like fashion: the post-step point of a given `G4Step` instance coincides with the pre-step point of the following `G4Step` instance. It is important to stress that when a step crosses a geometry boundary (i.e. the interface between two adjacent volumes) the post-step point is physically located on such boundary and logically belongs to the next volume.

A track represents the state of a particle but is not a collection of steps. Instead, the collection of particle states, i.e. the collection of `G4Track` instances, is referred to as an *event* in Geant4 and is represented by the `G4Event` class. This is the basic simulation unit and, at the beginning of an event:

1. primary track objects are generated and pushed to a Last-In-First-Out (LIFO) stack;
2. the last added track is popped from the track stack and simulated;
3. when secondary tracks are generated they are also pushed to the stack to proceed with the simulation.

A track is deleted under one of the following conditions:

- the track leaves the outermost volume, therefore leaving the simulation's universe;
- the particle described by the track undergoes a destructive interaction, such as a neutron capture;
- the particle's kinetic energy becomes zero and no alternative at-rest processes are available;
- other user-defined conditions that may belong to the physics lists implementation or to be motivated by other considerations.

A Geant4 event is terminated when the track stack is empty. The collection of Geant4 events is defined as a *run* and represented by the `G4Run` class.

4.1.6 Data extraction

Geant4 does not contain any embedded data analysis tool and only provides handles for the user to bind to in order to implement the data extraction logic. Intermediate information are exposed in sensitive detectors and in user actions.

Sensitive detectors are regular volumes for which the ability to react to particle interactions has been activated. At each step a particle does inside a sensitive detector volume, the `ProcessHits` method from the abstract `G4VSensitiveDetector` class is invoked. A pointer to the current `G4Step` is provided to the user.

User actions comprise a larger set of interactions points, at different level of detail, as summarized in Tab. 4.1

Table 4.1: Summary of the available User Actions in Geant4. A pointer to the current instance of a Geant4 object (run, event, track or step) is provided to the user at the corresponding condition.

Geant4 class name	Available handle	Condition
<code>G4UserRunAction</code>	<code>const G4Run*</code>	Begin and end of each run
<code>G4UserEventAction</code>	<code>const G4Event*</code>	Begin and end of each event
<code>G4UserTrackingAction</code>	<code>const G4Track*</code>	Begin and end of tracking of a <code>G4Track</code>
<code>G4UserSteppingAction</code>	<code>const G4Step*</code>	At each step
<code>G4UserStackingAction</code>	<code>const G4Track*</code>	At classification of a new <code>G4Track</code>

The user can choose which of the classes listed in Tab. 4.1 are more suitable for the simulation and implement them. Integration with data analysis tools, such as the ROOT Data Analysis Framework [128], can be conveniently implemented by expanding the aforementioned classes or, in more complex cases, by inheriting from them and take advantage of polymorphism. The only requirement for the data extraction to work is that the `G4RunManager` class is made aware of the classes the user intends to use.

4.2 Fundamentals of software optimization

One of the main objectives of this work is to address the performance of the Geant4 simulations in order to reduce their execution time and, therefore, maximize the use of the available resources. In this study, it is required that no modifications are introduced to the existing source code.

To accomplish this task, it is necessary to understand the general operations of a computer system: how memory is organized, how programs are translated into machine language instructions and how the workflow can be optimized. This is also necessary in order to set upper limits to the achievable improvements before a complete overhaul of the existing source code is required.

Although possible in principle, modifications to the simulation source code would require a large effort from many work groups and this is currently not planned. In addition, changes to the code require extensive validation against experimental data to ensure that the simulation results are not affected.

4.2.1 Abstraction in computer systems

A fundamental concept in computer systems, from hardware to software, is *abstraction*. For instance, several different technologies for storage (mechanical drives, solid state drives, tape, optical drives, etc.) exist, but from a programmer's perspective it is sufficient to know that they can all handle files in the same way, performing the same expected operations. Therefore, files represent an abstraction.

The same concept holds for software development: the behaviour of a given function is expected to remain the same regardless of the specific implementation as long as the interface, or the function's signature, remains unchanged.

A hierarchy of three fundamental abstractions can be identified, as shown in Fig. 4.2:

- Files

- Virtual memory
- Processes (and the extension to the concept of threads)

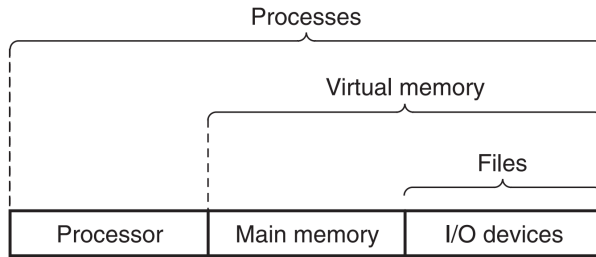


Figure 4.2: Fundamental abstraction hierarchy in a computer system [129].

4.2.1.1 Processes and threads

A process is the abstraction provided by the operating system for a running program. Multiple processes can (seemingly) run concurrently, i.e. they are in fact interleaved and continuously paused and resumed. This mechanism is known as *context switching*.

Multi-core processors are now widespread and, since they comprise physical replicas of the same logic units, they can actually run multiple processes concurrently.

Single-core CPUs can only run one process at a time and can handle concurrent execution by means of context switching; multi-core CPUs provide instead real concurrency. An intermediate concept is represented by *threads*, which can be seen as multiple execution units of a given process.

Unlike different processes, different threads share the same process' context and it is therefore easier to share data between the threads. From an instructions' execution perspective, multiple threads on a given CPU physical core allow the processing of seemingly concurrent instructions.

It is worth expanding on the concept of concurrency: while it generally refers to a system with multiple and simultaneous activities, the term *parallelism* conveys the idea of using concurrency to make a system perform its tasks faster. Three level of parallelism can be identified, from highest to lowest level:

- *Thread-level concurrency*: multiple programs can execute at once. Threads allow multiple control flows within each process. *Hyper-threading* (also known as simultaneous multi-threading) is a technology used by some CPUs to execute multiple control

flows at once and entails an intermediate design between a single-core and a multi-core CPU, in which only certain parts of the CPU hardware are duplicated, whereas others remain shared. As a reference, a conventional processor requires ~ 20000 clock cycles to shift between different threads, whereas a hyper-threaded CPU selects which of the threads to execute on a cycle-by-cycle basis. The choice is based on the projected time requested, for instance, to bring some missing data (*cache miss*) from disk or RAM into a closer cache: while waiting, the CPU can proceed with a different thread, therefore maximizing the use of the available resources.

- *Instruction-level parallelism*: up to 100 multiple instructions can be executed at once per each clock cycle.
- *Single-instruction, multiple-data* (SIMD) and *Single-instruction, multiple-threads* (SMT) parallelism: modern processors support single instructions triggering the execution of multiple operations in parallel, e.g. summing multiple pairs of single-precision floating-point numbers.

4.2.1.2 Virtual memory

Thanks to the virtual memory abstraction, each process can run as if it had exclusive control over the main memory. This uniform view of memory is referred to as *virtual address space* and a representation of its layout is given in Fig. 4.3.

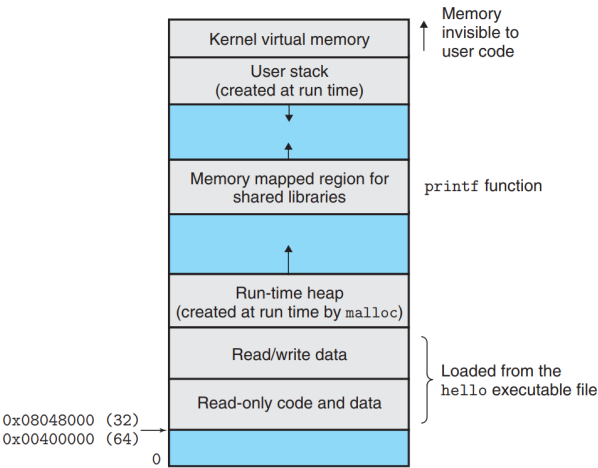


Figure 4.3: Representation of a process' virtual address space [129].

Virtual memory is structured as follows:

- Program code and data is the lowest-addressed read-only area, followed by the read/write portion of program data. They are both initialized from the contents of the executable object file.
- The code and data area is followed by the heap, which is defined at run time, e.g. by `malloc` or `new`, and is dynamically resized during the lifetime of the running process.
- The central portion of the virtual address space is reserved for the code and data of the shared libraries (shared libraries will be discussed in more detail in Sec. 4.2.2.2).
- On top of the heap (i.e. at higher memory addresses), memory is reserved for the user stack, which is also allocated dynamically at run time. The biggest difference between stack and heap memory addresses is visible from within the program's code in terms of the different scopes of the objects.
- The topmost section of the address space is not accessible to the program, as it is reserved to the operating system.

4.2.1.3 Main memory handling

As discussed, virtual memory is a convenient abstraction to ensure each program has a consistent and well-defined representation of the memory. Ultimately, the various areas of virtual address spaces form the main memory, usually referred to as *dynamic random access memory* (DRAM). Its function is to hold the programs and the data they manipulate while the processor is executing the various instructions. Logically, it is represented by a linear array of bytes, in which each location is identified by a positive and non-zero integer number, the memory address.

The processor (or CPU), the logic device responsible for executing the instructions stored in main memory, stores in its innermost storage locations, the *registers*, the memory addresses of the current instruction and data. More specifically, the instructions to be executed are pointed at by the *program counter* (PC); the CPU executes a given instruction and updates the PC to point to the next instruction. In fact, the set of operations provided by the CPU is very limited:

- *load*: overwrite a byte from main memory into a processor's register;
- *store*: overwrite a byte from a register into main memory;
- *operate*: perform an arithmetic operation between two values taken from two registers. The result overwrites the contents of the register used to store the result of the operation;

- *jump*: extract data from the instruction itself and copy it into the PC.

All data transfers in a computer happen as electrical signals sent over connections called *buses*. From the list above, it is clear that essentially all computer operations involve moving and copying data from one location to another and slow data transfers represent a strong limit to the computation speed.

The Dynamic Random-Access Memory (DRAM) only holds data for processing, whereas long-term storage is located further away from the processor. As a general rule, larger storage devices are slower to operate than smaller devices and also significantly cheaper to build. In order to perform fast calculations, data needs to be efficiently moved from long-term storage, into DRAM and then into the small and fast CPU registers. In order to fill the gap between the registers and the DRAM, CPUs are complemented with several layers of *cache memories*, serving as temporary staging areas for data likely to be needed in the near future by the CPU.

The full picture of the memory arrangement is given in Fig. 4.4, where the memory hierarchy is represented, from the smallest/fastest L0 level (the CPU registers) to the largest/slowest L6 level consisting of long-term storage tools, such as network file systems.

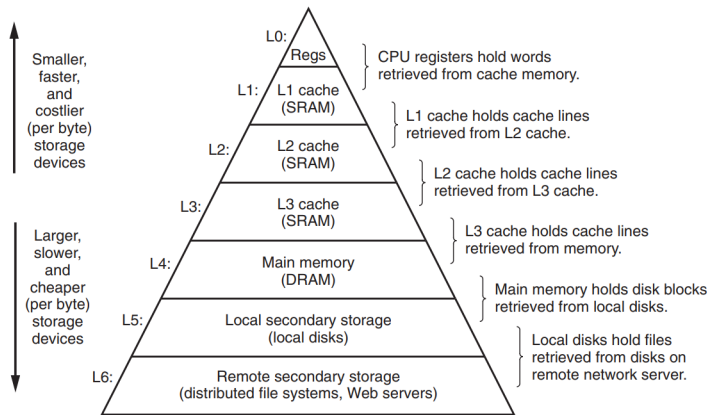


Figure 4.4: Representation of the memory hierarchy, from small and fast CPU registers (L0) to larger and slower network-connected long-term storage (L6) [129].

4.2.2 Compilers and linkers

Modern processors are capable of executing tens or even hundreds of instructions at the same time and an effective use of parallelism can significantly improve a program's performance. On the other hand, efficient management of the memory allows to minimize the number of required data transfers.

In the following sections, the details of translating a source code into an executable application will be discussed in detail. Computational and memory optimization options will also be analyzed.

Any high-level code, such as a C++ program, must be translated into low-level machine language instructions and packaged into an *executable object program* (or *executable object file*).

This task is carried out by other programs and, on Unix systems, a *compiler driver* like GCC is responsible for this. In Fig. 4.5 a simplified representation of a compiler driver's workflow is shown [129].

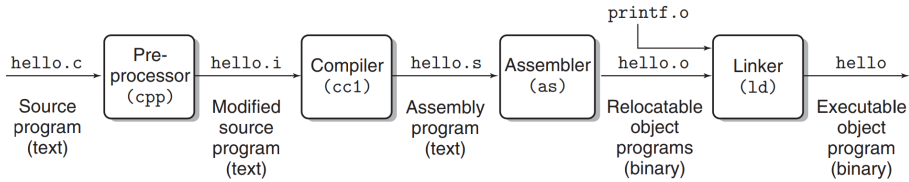


Figure 4.5: Representation of the compilation workflow: the preprocessor, compiler, assembler and linker stages are identified, together with the required input and the expected output [129].

The process is structured as follows:

- Preprocessing phase (cpp): all the directives starting with `#` are replaced with the corresponding contents, e.g. `#include` lines are replaced with the full header file's contents;
- Compilation phase (cc1): the source code is translated into an assembly-language program;
- Assembly phase (as): the assembly code is compiled into machine language instructions and a *relocatable object file* is produced;
- Linking phase (ld): the compiled program is merged with calls to other libraries and functions. An *executable object file* that can be directly loaded into memory and executed is produced.

4.2.2.1 Compile-time optimizations

Once a program is working as planned, i.e. yielding correct and reliable results, it is possible to take advantage of the compilers' ability to optimize it for the target machine in order to make it run faster. It is possible to identify three key points for an effective optimization:

- Careful choice of algorithms during the design phase: it is not possible to optimize below the minimum required latencies for data transfers and below the minimum required number of processor cycles for a given computation. For instance, different sorting algorithms scale differently with the number of elements to sort and bad design choices cannot be corrected by optimization.
- Compiler-friendly code that can easily be understood by the optimizer: the compiler can rely on a general model describing the target machine, which must be supported by the programmer's knowledge of the target architecture. It is important to remember that the compiler performs safe optimizations, i.e. those that do not affect the outcome and behaviour of the program, unless otherwise instructed. In order to do this, when the compiler cannot figure out the best strategy, it will resort to skip the optimization. One example of this behaviour is with conditional blocks (`if ... else` statements). Processors pre-execute code beyond the current instruction. When conditional blocks are found, a guess is made and data is calculated in advance. In case of mispredictions, a penalty is incurred. Although modern processors are capable of pre-executing all conditional branches in advance and only copy the correct one (conditional data transfer), it is still important to provide the compiler with predictable code. This should be taken into account especially when working with the GCC compiler, which tends to be less aggressive than others and to rely more on the programmer's good code [129]. The process of optimizing the code is not straightforward and compilers apply sophisticated algorithms to improve the quality of the code. It is usually sufficient to specify the high-level optimization flag, e.g. `-O2` or `-O3`, which triggers a consistent set of flags automatically. Level 2 has become the de-facto standard for most GCC use cases.
- Use of parallelism to ensure concurrency of control flows. This also implies writing parallel-oriented code, which can take advantage of multiple threads. Advanced libraries providing optimal inter-thread communication exist, but the benefits of parallelism are limited to the ability to express the required computations in assembly code. If the required degree of parallelism exceeds the number of available registers, the compiler *spills* and stores the extra values in the stack, i.e. at larger distances from the processor. As discussed earlier, this results in a performance drop due to the longer read/write times.

A good optimization is therefore the result of a trade-off between these three concepts, but it is nevertheless possible to identify some general guidelines that hold for a large variety of cases:

- Algorithms should be chosen according to the requirements of the design;

- The number of function calls should be minimized and computations should be moved outside loops;
- Excessive memory references should be avoided by preferring temporary variables to hold intermediate results; arrays and global variables should only be used to store final results;
- Loops should be unrolled;
- The use of instruction-level parallelism should be increased;
- Conditional data transfers should be facilitated.

The optimization techniques discussed until now are especially useful for relatively small projects. For larger codes, such as Geant4 or even the Athena framework, a higher-level approach should be considered in order to identify the most critical sections of the code. This is usually carried out by *profilers*.

Profiling entails running a fully-functional version of the program which has been compiled to include the so-called *instrumentation*. This diagnostic code allows to produce a list of all the functions called with the respective frequency together with an overview of how much time each part of the code takes to run. Indeed, it could be worth gaining a fractional improvement on a function called hundreds of times rather than a substantial improvement on a function only called a few times.

In Chapter 5, the effects of compile-time optimizations with pre-defined optimization levels on Geant4 simulation execution times will be investigated in detail.

4.2.2.2 Linking and build types

As shown in Fig. 4.5, once the compiler has translated the source code into relocatable object code, the linker is invoked. Linking is the process of weaving together all the pieces of the code and data into a single executable object file that can be loaded, i.e. copied to memory, and executed. In addition to relocatable object files (the output from the compiler) and executable object files (the output from the linker), also *shared object files* exist and they are special relocatable object files that can be loaded into memory dynamically.

From the linker's perspective, a software is composed of *symbols* and *symbol references*. Functions and initialized variables are called *strong symbols*, whereas uninitialized variables are referred to as *weak symbols*. The task of the linker is to resolve each symbol references found in the code by associating it with exactly one symbol definition from the symbol table provided by each input relocatable object file. Only three rules apply: multiple strong symbol

cannot coexist, if one strong and several weak symbols are found, the strong symbol is chosen and, lastly, if multiple weak symbols are identified, any of them can be preferred, which leads to unpredictability.

Once the linker has matched each symbol reference to exactly one symbol definition, *relocation* takes place. During this process, input modules are merged and absolute run-time addresses are assigned to each symbol. Relocation is divided in two steps:

- relocation of sections and symbol definitions: run-time memory addresses are assigned to the aggregated sections. At the end, each instruction and global variable has a unique run-time memory address;
- relocation of symbol references within sections: symbol references in the body of the code are updated to reflect the absolute addresses calculated at the previous step.

Linking can be done at compile time by the compiler (static linking), at load time by the loader (dynamic linking) or at run time by the program itself (dynamic linking). Different link methods affect the application's running times, due to the different memory access schemas. The effects of dynamic and static builds on Geant4 simulations are analyzed in Chapter 5.

Static linking Static linking is the most straightforward technique to merge the various relocatable object files.

Static libraries are collections (*archives*) of relocatable object code. This means that the source code has already been compiled into machine language, but absolute references have not been calculated. When a user call is made, for instance, to a function from the C++ standard library, the linker retrieves the appropriate archive and only copies the binary code relative to the referenced function into the executable file.

This approach has several advantages, because it enforces some of the general optimization guidelines: jumps and references are minimized, unnecessary copies are avoided. Being part of the code and data segments, statically linked libraries can be copied into the RAM at a very early stage and can reside closer to the processor.

However, for the same reason of being embedded into the program's code, static libraries result in larger executables, which in turn means penalties in memory consumption, increased storage footprint on disks and longer transfer times over slower buses [130]. In addition, the explicit dependence on a specific version of a static library means that any change to the library itself can only be deployed with the re-linking of all the depending applications.

Dynamic linking To reduce the executable disk footprint and to facilitate the maintenance of libraries and code, *shared libraries* (or shared object files) have been introduced. They can be loaded either at load time or at run time and they can be copied to any arbitrary memory address by the dynamic linker.

At compile time, a partial static link is carried out, but it is only finalized when the program is started. No data from the shared object file is copied into the executable object file. Instead, the linker copies some relocation and symbol table information that will allow to resolve memory addresses at run time.

When an executable is started, control is transferred to the *loader*. Firstly, all required shared objects are relocated into some available memory address. Then, the symbol references in the executable are relocated to point to the symbols defined by the shared object. At the end of this process, the loader transfers control to the application and memory addresses are not allowed to change anymore.

Lastly, it is worth pointing out that dynamic libraries can also be explicitly loaded and unloaded at run time.

Position Independent Code (PIC) Position Independent Code (PIC) allows flexibility in the relocation process: if dynamic libraries were assigned immutable memory addresses, memory would end up completely fragmented and unusable.

PIC code relies on the fact that the data segment is always allocated immediately after the code segment. This means that the distance in terms of memory addresses between a given instruction in the code segment and the data it manipulates in the data segment is a run-time constant. It is therefore sufficient to introduce an overall offset and recalculate absolute memory addresses based on the relative distances defined in the object code. Logically, this operation consists of redefining a new origin in a translated reference frame.

The compiler creates a *global offset table* (GOT) at the beginning of the data segment plus a relocation record for each entry in the GOT. At load time, the dynamic linker relocates each entry in the GOT and produces absolute memory addresses.

PIC is an effective way to maximize memory occupancy, but from a computational standpoint it is very intensive: each global variable reference requires five instructions to be loaded plus an additional memory reference to the GOT. Moreover, PIC uses an additional register to hold the address of the GOT entry. For register-poor processors, this results in the spilling of the registers onto the stack, with substantial performance degradation.

To mitigate the effect of the extra instructions, Executable and Linkable Format (ELF) compilation systems, such as Linux, use a technique called *lazy binding*: the matching of procedure addresses is deferred until the first time a procedure is called. The run-time

overhead is non-trivial only for the first call. The linking in this case relies on the interaction between the GOT and the Procedure Linkage Table (PLT), which after pointing at the procedure's memory address rewrites the appropriate GOT entry in order to ensure that subsequent calls immediately transfer control to the called function [129].

As it will be discussed in Chapter 5, PIC is enabled by default when building Geant4 libraries as shared objects.

Chapter 5

Performance optimization of simulation workflows

ATLAS has a globally distributed software project with millions of lines of codes. The large amount of data and the complexity of the code require workflows for data reduction and extensive software optimizations. To interpret the data correctly, interesting physics interactions as well as background are simulated as collision events in the detector. Due to the fine granularity of the ATLAS detector, millions of events must be simulated. Simulations take up to 40% of the total CPU load time. It is therefore important to achieve an improvement in speed that can have significant effect on the precision of physics analysis: more statistics would allow for better modeling of signal and background; moreover, optimized performance and better exploitation of the currently available hardware would postpone the need for new investments.

A possible solution for running faster jobs using less CPU resources is fine tuning the compiler's optimization options. In general, with no optimizations, the compiler's goal is to reduce the cost of compilation and to make debugging produce the expected results; on the contrary, turning on optimization makes the compiler attempt to improve the performance and/or code footprint at the expense of compilation time and possibly the ability to debug the program.

In this chapter, the effects of different optimization options and different build types are presented in order to identify viable options to reduce the execution time of the full Geant4 simulations.

5.1 Methodology

5.1.1 Details of the Geant4 benchmark simulation

The studies on the simulations' execution times have been carried out using a stripped-down Geant4 benchmark simulation [131]. The simulation only comprises the fundamental components:

- Random number engine;
- Physics list and magnetic field;
- Particle source definition;
- Simple measurement of deposited energy.

The code does not provide an embedded geometry definition, but is instead necessary to import a GDML file containing the geometry and material specifications (see Sec. 4.1.3).

The Mersenne Twister pseudo-random number generator (PRNG) is adopted by the benchmark simulation. The algorithm, first proposed in 1998, is characterized by a period of $2^{19937} - 1$ and is widely used in many open source and commercial applications [132]. The implementation used by the simulation is provided by the CLHEP library [133].

The FTFP_BERT physics list is chosen as the default option for the simulation. This physics list is built around the Fritiof model and the Bertini intra-nuclear cascade model [134–136]. The Fritiof model treats all hadron-hadron interactions as binary interactions, where the product states are excited states of the hadrons with either continuous or discrete mass spectra. The excited hadrons are considered as QCD-strings, and the corresponding Lund-string fragmentation model is applied in order to simulate their decays (see Chapter 1) [135]. A 2 T magnetic field is applied to the detector, which corresponds to the nominal magnetic field in the central solenoid¹.

The Geant4 General Particle Source (GPS) is used as the particle generator [137]. The primary proton-proton collision is not simulated and post-hadronization products are instead used as primary particles. In particular, investigations using protons and positive/negative

¹The magnetic field used for the simulations presented in this thesis is an extreme simplification of the actual ATLAS magnetic fields, which consist of: a) a solenoid surrounding the Inner Detector and providing a 2 T axial magnetic field; b) three toroid magnets providing the magnetic field for the muon spectrometer with the following properties: 1.5 to 5.5 T at $0 < |\eta| < 1.4$ and 1 to 7.5 T at $1.6 < |\eta| < 2.7$.

pions have been carried out. Additionally, pure geometry tests with geantinos² have also been performed. Generation of randomized primary particle types is also supported by the simulation code.

Three different energies, namely 10, 20 and 50 GeV, have been chosen in order to stress the simulation with computationally intensive particle interactions while still being representative of the actual ATLAS operating conditions. The typical energy fractions of different particles as a function of their energy are shown in Fig. 5.1.

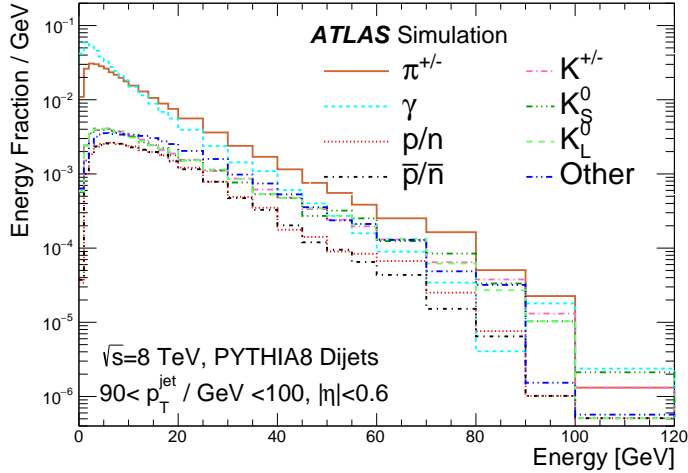


Figure 5.1: Fractional contribution to the total jet energy of particles with different momenta [139].

As discussed before, the benchmark simulation provides minimal data analysis functionalities and only collects the total deposited energy. The stepping action is responsible for the data acquisition and the total energy deposit for any particle in the whole Geant4 world volume is collected at each step. To ensure complete decoupling between the user-provided geometry description and the general purpose benchmark code, no details of the energy deposition in any detector part are stored. At the end of each event, the total energy deposited is retrieved from the stepping action class and the run action class is updated accordingly. At the end of the run, the total and average (per event) energy deposits are calculated and printed to screen.

Lastly, the execution time is automatically monitored by the simulation by means of the `G4Timer` class. Several factors can affect the execution time of a computer process. Indeed,

²Geantinos are chargeless, massless, completely non-interacting particles which can be used for geometry and tracking diagnostics. A ChargedGeantino also exists and allows proper tracking in a magnetic field. Geantinos only support transportation and cannot be assigned any interaction processes [138].

as discussed in Chapter 4, data is to be retrieved from a complex hierarchy of memory levels and, especially for larger objects such as physics lists, the overhead generated when copying the data from disk to RAM and then to the CPU registers can be significant. Therefore, a reliable execution time measurement must not include these system-dependent steps.

The Geant4 simulation used for this work calculates the net execution time of a Geant4 run by starting a timer inside the `BeginOfRun` method and stopping it inside the `EndOfRun` method. The standard Unix `time` data is returned, subdivided in User, Real and System. The analysis presented in this thesis only takes into account the effective wall time (Real).

Limiting the time measurements to the extent of a single run minimizes the effects of disk or network speeds, but it is nevertheless extremely important to ensure that the processing time is not affected by concurrent processes running on the benchmark machine, which can result in unpredictable sharing of the common resources. For this purpose, a dedicated standalone machine at CERN, not shared with other users, has been used. In addition, simulations were also performed on the Lund University Aurora cluster; in this case, the requests for node allocations have been complemented with the requirement of exclusivity. One Aurora node consists of two 10-core CPUs; an exclusive allocation of a single node entails that, regardless of the number of concurrent processes used by the simulation, the full node is reserved for the simulation and all the resources are exclusively allocated to the requesting user. A summary of the characteristics of both the machines used is presented in Tab. 5.1.

Table 5.1: Summary of the computing resources.

	CERN standalone machine	Compute node on Lund University cluster
CPU	2× Intel Xeon E5-2630 v3 2.40GHz	2× Intel Xeon E5-2650 v3 2.30GHz
Architecture	64 bit Haswell x86_64	64 bit Haswell x86_64
N. of cores	16	20
Threads per core	2	1
Cache	20 MB (L1: 64 KB, L2: 256 KB, L3: 20 MB)	25 MB (L1: 64 KB, L2: 256 KB, L3: 25 MB)
RAM	64 GB	128 GB
Filesystem	XFS	IBM General Parallel File System (GPFS)
Operating system	CentOS 7	CentOS 7

5.1.2 Runtime geometry representations

Geant4 provides an advanced interface to describe the geometry for the simulation. Most solids can be represented with dedicated classes, such as `G4Box` for boxes or `G4Tubs` for cylindrical solids, or with general-purpose solids capable of representing arbitrary shapes. However, defining volumes and materials with this procedure is very labour-intensive, es-

pecially for modeling complex geometries.

To facilitate the definition of large detectors, such as those in operation at LHC, geometries produced outside Geant4 can be imported and actively handled from within the simulation code. In addition to allowing cross-operation between CAD software and Geant4, modular geometries loadable at runtime facilitate the maintenance of the detector descriptions, as they can easily be replaced without the need to recompile large portions of the code.

The two alternatives used for this thesis, namely GDML and GeoModel-based geometries, are discussed in detail.

5.1.2.1 The GDML format

The Geometry Description Markup Language (GDML) is a specialized, application-independent, geometry description XML-based format. The GDML format is suitable both as the primary geometry implementation language as well as a convenient data exchange format for the existing applications [140, 141].

GDML files are structured in blocks identified by `< >`-delimited tags. At the highest level, five elements can be defined (see Fig. 5.2 for an example of a simplified GDML file):

- `<define>`: this block allows the definition of global named constants, quantities, expressions, positions and rotations to be used in the scope of the document;
- `<materials>`: this block is reserved to the definition of the general material properties. Elements and compounds can be cross-referenced and composed to obtain more complex materials;
- `<solids>`: similarly to the materials block, arbitrarily complex solids can be obtained from the composition of simpler objects. The most fundamental shapes must be declared before the more complex ones;
- `<structure>`: in this block, solids and materials are matched and the hierarchy of the detector volumes is defined. Optionally, additional GDML files can be referenced and imported in this block with the `<file>` tag. This improves the modularity of the full detector description and improves the readability and maintainability of the GDML code;
- `<setup>`: as required by Geant4, the outermost volume for the simulation, i.e. the *world* volume, is declared in this block.

Geant4 provides a built-in GDML parser which can optionally be built together with the standard Geant4 libraries. The parser acts as an interface between the XML code and the



Figure 5.2: Layout of a simple GDML geometry file [141].

three fundamental Geant4 stores for geometry objects: the solid, logic and physical volume stores. Parsed objects are first translated into the corresponding Geant4 classes and then directly sent to the stores. During the detector construction phase, after the GDML file is successfully imported, pointers to the instances of the stored volumes can be retrieved for further processing.

An overview of the structure of the ATLAS detector as represented by the GDML geometry file is given in Fig. 5.3. Note that this geometry does not contain a definition for the End Cap Electromagnetic Calorimeter (EMEC) as the format is lacking support for the required shapes.

5.1.2.2 The GeoModel toolkit

A different approach to overcome the difficulties of detector geometry description is represented by the GeoModel toolkit. It consists of a library of geometrical primitives which can be used to describe detector geometries. The toolkit is intended especially for large and complex detectors and is optimized for minimum memory footprint [142, 143].

The GeoModel is a C++ package used to build a graph of geometrical primitives composing a scene graph³. A detector's geometry description consists of a tree of GeoModel nodes

³A scene graph is a structured tree of nodes representing graphical primitives with the relative appearance properties, organized to produce the final image.

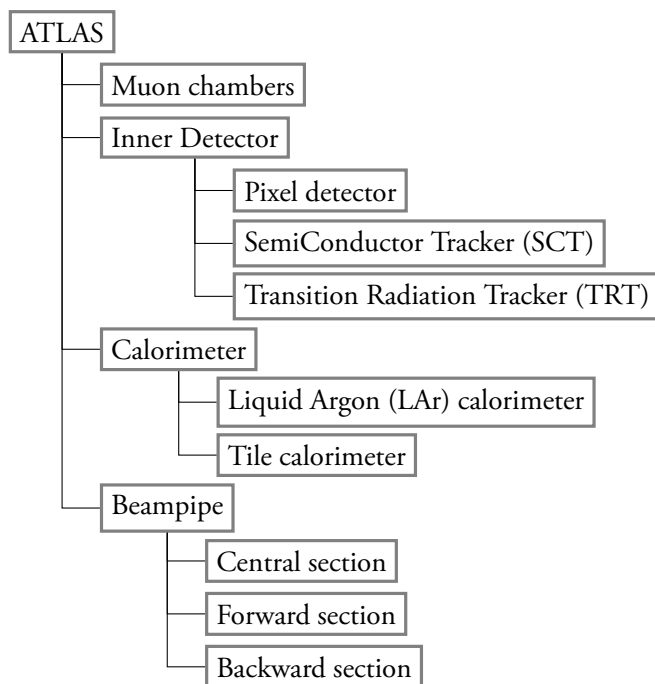


Figure 5.3: Structure of the GDML representation of the ATLAS detector.

resembling a tree of volumes. The geometry graph comprises both volumes positioned in space as well as parameterized volumes, which can be combined to form complex shapes.

GeoModel provides a limited number of geometry primitives and instead relies on boolean shapes, obtained by unions and intersections of simpler objects. In addition to these, volumes can also be parameterized; this entails definition of prototype volumes together with mathematical objects defining replicas and parameterized positions. This algorithmic representation of a detector's geometry can be conveniently packaged into a shared object library and dynamically linked at runtime.

Once the GeoModel tree is executed and inflated into a full geometry description, the GeoModel-to-Geant4 interface can store the objects into the solids and logical/physical volumes Geant4 stores, similarly to what is done for the handling of GDML files.

5.1.3 Compiler's optimization levels

As discussed in Chapter 4, the compilation process is responsible for translating source code into machine language instructions. This process can be subject to several stages of optimization in order to make it, for instance, smaller in terms of storage and faster in terms

of execution time.

Many flags are available for enabling very specific optimizations aimed at improving the performance and/or code size, usually at the expense of compilation time. When the compiler is invoked, a list of individual flags can be passed to the compiler command line, but it is often more convenient to take advantage of the predefined grouping of such flags into *optimization levels*, activated with the flag `-O` followed by the level identifier. Available identifiers are `0`, `1`, `2`, `3`, `s`, `g` and `fast`. The attention is hereby focused on the increasingly aggressive optimization levels `-Os`, `-O1`, `-O2` and `-O3`. Starting from a minimum set of flags in `-O1`, higher levels introduce additional options to the preceding level they build upon.

Excluding the default `-O0` option, which does not introduce any change, the first level `-O1` is responsible for eliminating unused sections of the code, optimizing conditional branches and improving the instruction execution order to minimize latency from cache misses, i.e. the time required to bring missing data to RAM and to caches closer to the CPU (Sec. 4.2.1.3).

The `-O2` level inherits all of the properties of `-O1` and introduces additional steps:

- redefinition of constant values is propagated throughout the code;
- expressions common to multiple branches of the code are suppressed and moved to a single location;
- unused variables resulting from the redefinition of the constants are removed.

Loop unrolling is introduced at this stage and this is intended to facilitate parallel execution when supported by the hardware. Inlining is applied for simple functions called multiple times.

In general, `-O2` enables all optimizations that do not involve a size-performance tradeoff. The gain in efficiency is achieved at the expense of compilation time.

The highest optimization level `-O3` is the most aggressive and, for badly designed codes, it can result in unexpected changes of the program's functionality. Prefetching, i.e. the retrieval of data before they are used to reduce latency due to cache misses, is enabled and floating point instructions are transformed. For instance, floating point divisions can require 20 to 60 CPU clock cycles; instead, if they are transformed into multiplications, only 5 to 10 clock cycles are necessary.

Lastly, `-Os` is an optimization level aiming at minimizing the compiled code storage size and is not expected to introduce performance-oriented optimizations.

Additional options exist, such as `-Ofast`, which introduces unsafe math optimizations, and `-Og`, which optimizes for debugging. However, these are outside the scope of the studies presented in this thesis and will not be further investigated.

5.2 Impact on the simulation execution time

In this section, the effects of build types, optimization flags and primary particles on the Geant4 execution times are investigated.

5.2.1 Studies with different build types

In Sec. 5.2.1.1, preliminary studies on the ATLAS Inner Detector geometry are presented. The impact of static and dynamic build types on Geant4 runtimes is evaluated.

Further tests using a more complete version of the GDML geometry (without EMEC) are presented in Sec. 5.2.1.2. In addition to the default static and dynamic builds (i.e. consisting of a set of multiple shared objects), a new packaging for the dynamic build, resulting in a single shared library file, has been tested [144].

5.2.1.1 Comparison between static and dynamic build

To evaluate the effects of different build types on the execution time, both static and dynamic Geant4 libraries have been produced. The standalone Geant4 simulation described in Sec. 5.1.1 compiled both statically and dynamically against Geant4 v10.5.0 has been used as benchmark. As previously discussed, this simulation is based on GDML geometries loaded at runtime, rather than hard-coded detector definitions.

The relevant CMake options used for the dynamic Geant4 build are listed:

```
-DGEANT4_BUILD_MULTITHREADED=ON \  
-DGEANT4_USE_GDML=ON \  
-DBUILD_SHARED_LIBS=ON \  
-DCMAKE_CXX_FLAGS="$CXXFLAGS -g -O2" \  
-DCMAKE_C_FLAGS="$CFLAGS -g -O2" \  

```

The corresponding options for the static build are:

```
-DGEANT4_BUILD_MULTITHREADED=ON \  

```

```

-DGEANT4_USE_GDML=ON \
-DCMAKE_CXX_FLAGS="$CXXFLAGS -g -O2" \
-DCMAKE_C_FLAGS="$CFLAGS -g -O2" \
-DBUILD_SHARED_LIBS=OFF \
-DBUILD_STATIC_LIBS=ON \

```

For this preliminary study, an isotropic source centered at (0,0,0) has been used and 5000 pions characterized by an energy of 50 GeV have been considered as source particles.

Three versions of the GCC compiler, namely 4.8.5, 6.2.0 and 8.2.0, have been used for these investigations and a comparison between four GCC optimization flags (0s, 01, 02 and 03) has also been performed. The default optimization level used by most build systems is 02, hence in this study it has been considered as a reference.

For each of the studied configurations the benchmark simulation has been run 5 times. Average values are presented and for all the studied configurations, standard deviations are of the order of 2%.

The computations have been carried out on a standalone machine at CERN and on a university cluster in Lund (see Tab. 5.1). On both machines, CPU and memory resources have been exclusively allocated for these simulations to prevent uncontrolled overhead from concurrent processes (other than the minimum operating system's tasks). CPU multithreading on the university cluster nodes is disabled.

In Figs. 5.4 and 5.5, the comparison between static and dynamic performance using the ATLAS inner detector geometry is given. In Fig. 5.4, the results of the computations carried out on the cluster are presented: with the static approach the execution time can be scaled down by 10%. Fig. 5.5 shows the results for the calculations performed on the standalone CERN machine. The static approach, for all GCC versions, reduces the execution time by more than 9%.

Fig. 5.6 shows the trend of the execution time as a function of the number of threads used for the computations. The number of primaries per thread has been kept constant (1250 primaries/thread) for all the cases. Therefore, for a doubled number of threads, the number of primaries is also doubled; the effect of parallelization ensures that the resulting execution time is not doubled, despite a minor increase of the order of 10% between 1 and 16/20 threads, which can be ascribed to CPU and I/O overhead. The improvement between static and dynamic linking is confirmed in all the analyzed cases. CPU multithreading, i.e. the virtualization of CPU threads, is responsible for the higher increase in execution time observed for 20 and 32 threads on the standalone machine, which has only 16 physical cores. The same effect is not observed on the Lund cluster because no multithreading is available and only 20 physical cores can be used for the computations.

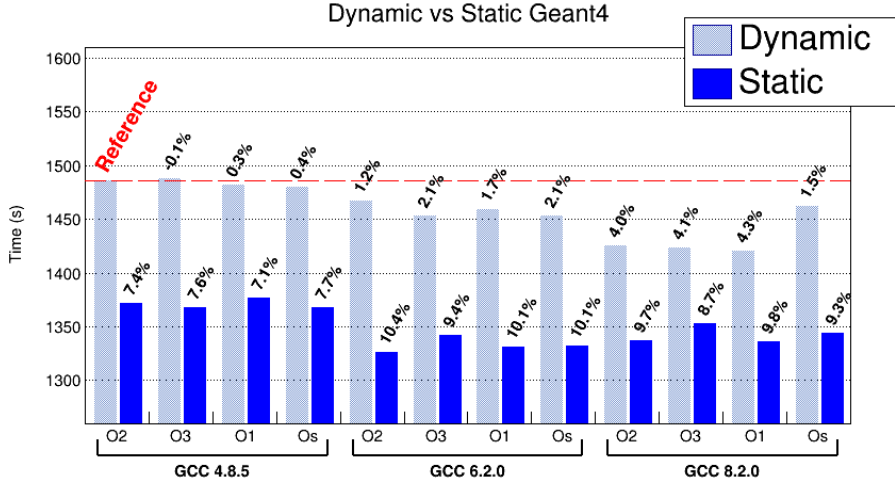


Figure 5.4: Comparison between dynamic and static performance with inner geometry (percentages indicate the speedup with respect to the reference case). The computations have been carried out on Lund cluster. For the computations 50000 primary particles have been considered and 4 threads have been used. Note that the time axis starts at 1250 s.

5.2.1.2 Geant4 as a single library

In this section the impact of using a single dynamic library will be evaluated, as compared to the previously discussed multi-library dynamic and static builds.

For these tests, a GDML geometry comprising the full inner detector, the LAr hadronic and tile calorimeters, the EM barrel and the muon spectrometer has been used. This geometry does not contain a definition for the electromagnetic calorimeter end-cap (EMEC). As primary particles, 2500 50 GeV negative pions have been used.

The code was built against Geant4 10.5.0 on a CERN standalone machine (see Tab. 5.1) with GCC 6.2.0 and 8.2.0 compiler versions, four optimization levels and was executed with 4 threads.

To build the single dynamic Geant4 library for these tests, the CMake structure has been modified. The new flag `BUILD_SINGLE_LIB` was added; it is an optional flag and it must be enabled in addition to the standard `BUILD_SHARED_LIBS` and `BUILD_STATIC_LIBS` flags. This allows the choice of which build type should be used for the single library [144]. The optimizer options discussed earlier are unchanged.

Fig. 5.7 shows the results obtained with three different build types: static, dynamic (default

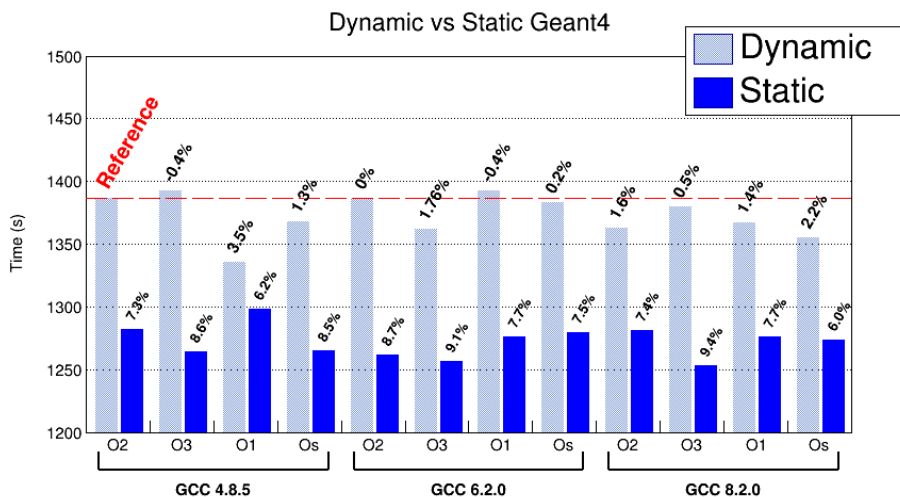


Figure 5.5: Comparison between dynamic and static performance with inner geometry (percentages indicate the speedup with respect to the reference case). The computations have been carried out on CERN standalone machine. For the computations 50000 primary particles have been considered and 4 threads have been used. Note that the time axis starts at 1200 s.

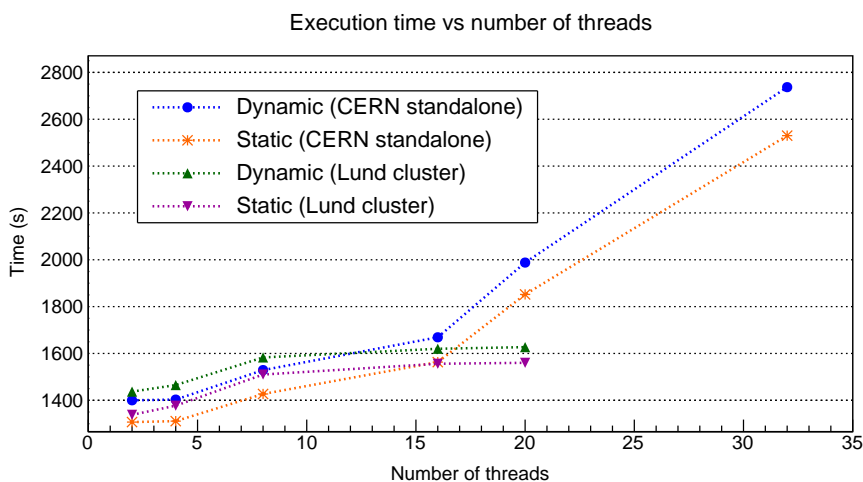


Figure 5.6: Execution time as a function of number of threads with inner detector geometry. The number of primaries per thread has been set equal to 1250. GCC 8.2.0 compiler has been used for the computations. Note that the time axis starts at 1300 s.

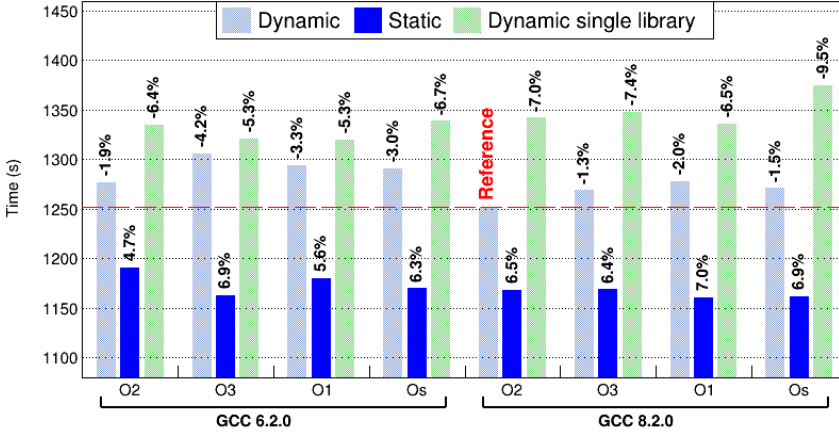


Figure 5.7: Comparison of execution times for a standalone Geant4 v10.5.0 simulation. For all cases the ATLAS GDML geometry description without the electromagnetic end-cap calorimeter has been used with an isotropic 50 GeV π^- source centered at (0, 0, 0). Differences in performance are expressed as a relative percentage with respect to the reference case: multi-library, GCC 8.2.0 with -O2 optimization. For each of the studied configurations, the benchmark simulation has been run 5 times. Average values are presented and in all cases, standard deviations are of the order of 2%. Simulations have been carried out on a standalone CentOS 7 machine, with an Intel Xeon E5-2630 v3 2.40 GHz, equipped with 64 GB of RAM.

multi-library configuration) and single dynamic library. For all cases, differences in performance are expressed as a relative percentage with respect to the reference case: multi-library, GCC 8.2.0 with -O2 optimization. For each of the studied configurations, the benchmark simulation has been run 5 times. Average values are presented and in all cases, standard deviations are of the order of 2%.

For both compiler versions, the single library approach exhibits an increase of $\sim 10\%$ in execution time. This effect seems to be counter-intuitive, but could be explained by considering how shared libraries call and load objects in memory and how the interaction between Geant4 core libraries and user application is structured. Each call to a function in a dynamic library takes advantage of a trampoline which reads the memory address of the called method from a lookup table and passes it to the calling function (see Sec. 4.2.2.2). This results in an increased number of calls and jumps, which eventually slows down the simulation execution [145].

5.2.2 Investigations on the effects of different optimization flags

Further investigations have been carried out to understand the effect of different optimization flags, which is not apparent from the results presented in the previous sections. An in-depth review of the available GCC optimization levels was given in Sec. 5.1.3.

It has been conclusively found that the default build mode for Geant4 (labelled as RELEASE) enables the following C++ compiler flags [146]:

```
CXX_FLAGS = -W -Wall -pedantic -Wno-non-virtual-dtor -Wshadow
            -Wno-long-long -Wwrite-strings -Wpointer-arith
            -Woverloaded-virtual -Wno-variadic-macros
            -pipe -DG4USE_STD11 -pthread -ftls-model=initial-exec

            -O3 -DNDEBUG -fno-trapping-math
            -ftree-vectorize -fno-math-errno

            -fPIC
```

The optimizer flag `-O3` is therefore always enabled and overwrites any different specification. In addition to the `-O3` flag, the options `-DNDEBUG` `-fno-trapping-math` `-ftree-vectorize` `-fno-math-errno` are also turned on.

In order to control the actual optimizer flag used during the compilation process, it is necessary to overwrite the default `_RELEASE` flags by setting appropriate CMake options:

```
-DCMAKE_BUILD_TYPE=Release \
-DCMAKE_CXX_FLAGS_RELEASE="-O2" \
-DCMAKE_C_FLAGS_RELEASE="-O2" \
```

This ensures that the applied set of flags is the following:

```
CXX_FLAGS = -W -Wall -pedantic -Wno-non-virtual-dtor -Wshadow
            -Wno-long-long -Wwrite-strings -Wpointer-arith
            -Woverloaded-virtual -Wno-variadic-macros
            -pipe -DG4USE_STD11 -pthread -ftls-model=initial-exec

            -O2

            -fPIC
```

Differences emerge once the new flags are adopted, both for compiling the Geant4 libraries and for building the simulation code, and their impact on the execution time is summarized in Tab. 5.2.

Table 5.2: Execution time for different optimization flags, specified according to the indications discussed in Sec. 5.2.2. The standard deviations are of the order of 1.5%.

Build type	Optimization level							
	-O2	Diff. (%)	-O3	Diff. (%)	-O1	Diff. (%)	-Os	Diff. (%)
Dynamic	1199 s	Reference	1218 s	-1.6	1315 s	-9.7	1474 s	-22.9
Static	1088 s	Reference	1065 s	2.1	1189 s	-9.3	1214 s	-11.6

5.2.3 Different primary particles

To estimate the impact of different particles on the simulation execution time, a first preliminary study has been carried out using the same GDML geometry; protons, positive/negative pions and geantinos, the massless virtual particles available in Geant4, were chosen as primary particles. For each of them, two energies were considered: 10 and 20 GeV and for each run 5000 primaries were generated (Tab. 5.3). All simulations were performed on the Aurora cluster at Lund University and full nodes were reserved with the `exclusive` option (see Tab. 5.1). The code was built against Geant4 10.6.2 and GCC 8.2.0.

In order to include the effect of the EMEC on the simulation execution time, support for the `GeoModel` representation of the ATLAS geometry has been added to the standalone simulation [141, 143]. The impact of different primary particles, namely charged pions and protons, at different energies (10, 20 and 50 GeV) has been evaluated (Tab. 5.4). Simulations have been run on a CERN standalone machine (Tab. 5.1) and built with GCC 8.2.0 against Geant4 10.6.2.

For all the primary particles analyzed, a decrease in the simulation execution time is observed for the static build, when compared to the dynamic case. This improvement is increasingly pronounced as the complexity of the interactions grows. Considering the geantinos⁴, a 5% decrease in time was observed (Tab. 5.3). The speed-up rises to more than 6% in the case of 50 GeV protons tested with the full ATLAS geometry (Tab. 5.4) and exceeds 10% in case of 20 GeV protons tested with GDML geometry (Tab. 5.3).

⁴Geantinos are massless virtual particles available in Geant4 as a probe to test ray-tracing propagation through the detector geometry. As such, they do not undergo any interaction and, in all cases, the observed execution time is ~ 3 s per run.

Table 5.3: Execution times per run for p , π^\pm and geantinos at 10, 20 and 50 GeV, tested with static and dynamic Geant4 builds. The GDML geometry (without EMEC) is used.

Particle type	Simulation time (s)	Decrease w.r.t. protons (%)	Increase w.r.t. static case (%)
Dynamic library (10 GeV, 20 runs)			
p	561 ± 9	—	10.8
π^-	554 ± 10	1.2	11.2
π^+	537 ± 5	4.2	10.5
Geantino	3.0 ± 0.1	1.99×10^4	5.6
Static library (10 GeV, 20 runs)			
p	506 ± 6	—	—
π^-	498 ± 8	1.5	—
π^+	486 ± 4	3.9	—
Geantino	3.2 ± 0.1	1.7×10^4	—
Dynamic library (20 GeV, 100 runs)			
p	1090 ± 14	—	10.8
π^-	1043 ± 19	4.3	9.8
π^+	1039 ± 18	4.7	10.2
Geantino	3.1 ± 0.1	3.64×10^4	4.7
Static library (20 GeV, 100 runs)			
p	983 ± 12	—	—
π^-	950 ± 14	3.3	—
π^+	943 ± 12	4	—
Geantino	3.0 ± 0.1	3.4×10^4	—

The mean simulation execution time of the positive and negative pions are consistent within the parameters' error.

Pions are mesons characterized by a rest mass $m_0 = 139.57 \text{ MeV}/c^2$; π^+ consists of the valence quarks $u\bar{d}$ and π^- of the valence quarks $d\bar{u}$. Since they are heavy charged particles, they lose energy through ionization processes described by the Bethe-Bloch formula; pions also undergo inelastic nuclear collisions leading to the creation of showers in calorimeters. These two types of pions are characterized by cross sections that are compatible (both around 25 mb) [15]: this justifies the compatibility in the execution times.

For energies of 20 and 50 GeV, computations with protons show a slightly longer execution time. Protons are baryons composed of three valence quarks (uud) with mass $m_0 = 938.27 \text{ MeV}/c^2$. Similarly to what happens for pions, protons undergo Coulomb interactions with the atomic electrons and inelastic collisions with the nucleons of the mate-

Table 5.4: Comparison of execution times for a standalone Geant4 simulation built with GCC 8.2.0. Four different isotropic sources, centered at $(0, 0, 0)$, of p , π^\pm and geantinos have been used, each tested with 10, 20 and 50 GeV primary energy. For all cases the full ATLAS GeoModel geometry description has been adopted. Simulations have been carried out on a standalone CentOS 7 machine, with an Intel Xeon E5-2630 v3 2.40 GHz, equipped with 64 GB of RAM.

Particle	Simulation time (s)	Decrease w.r.t. protons (%)	Increase w.r.t. static case (%)
Dynamic library (10 GeV, 45 runs)			
p	607 ± 9	—	6.7
π^-	617 ± 9	-1.6	6.7
π^+	600 ± 10	1.1	7.5
Static library (10 GeV, 45 runs)			
p	569 ± 10	—	—
π^-	578 ± 9	-1.6	—
π^+	558 ± 7	1.9	—
Dynamic library (20 GeV, 45 runs)			
p	1172 ± 13	—	6.7
π^-	1169 ± 18	0.2	6.9
π^+	1141 ± 14	2.6	6.3
Static library (20 GeV, 45 runs)			
p	1098 ± 11	—	—
π^-	1093 ± 14	0.4	—
π^+	1073 ± 13	2.3	—
Dynamic library (50 GeV, 45 runs)			
p	2757 ± 46	—	6.2
π^-	2712 ± 40	1.6	7.0
π^+	2675 ± 39	3.0	7.0
Static library (50 GeV, 45 runs)			
p	2596 ± 30	—	—
π^-	2533 ± 28	2.4	—
π^+	2499 ± 37	3.7	—

rial. At the same energy, protons are characterized by larger cross sections than pions [147]; at the energies considered in this study, the cross section is around 40 mb.

Since protons and pions are charged hadrons, in the presence of a magnetic field, they experience the Lorentz force and their trajectory is deflected according to their momentum

at the same energy. Considering the mass-energy equation $E^2 = (\mathbf{p} \cdot c)^2 + E_0^2$ (where $E_0 = m_0 c^2$), for particles at the same energy, the larger the mass the smaller is the momentum of the particle and, if the value of the p_T decreases, the radius of the curvature induced by the magnetic field will decrease. This difference in trajectory's radius could, in principle, contribute to a difference in the measured times. At the energies tested, however, the momenta of protons and pions differ slightly; this explains why no tangible discrepancies were found in the execution time of the simulation due to the difference in mass and trajectories traveled between the primary particles analyzed.

Showers induced by pions and protons inside calorimeters are characterized by different developments. Based on the baryon number conservation laws, it is more likely for proton-induced showers to evolve as hadronic showers and for pion-induced ones to develop as EM showers.

In addition, considering the fact that $X_0 < \lambda_I$, together with the large number of possible final states for particles in hadronic showers, it is reasonable to conclude that the simulation of an EM shower is less resource-intensive from a computational perspective.

However, as shown in Fig. 5.8, the response of the calorimeter to charged hadrons as described by the FTFP_BERT is comparable for protons and pions [139]. Hence, the differences in the observed execution times must be ascribed to the contribution from ionization losses inside the Inner Detector.

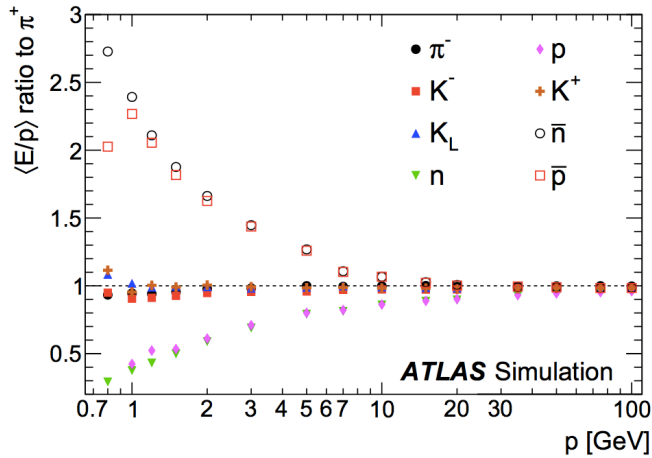


Figure 5.8: Calorimeter response to single particles of different species expressed as the ratio to the π^+ calorimeter response. The FTFP_BERT physics list is used [139].

5.3 Conclusions and outlook

5.3.1 Build types

Execution time for simulations based on Geant4 can be significantly improved by changing the default build method: a tangible improvement can be achieved by considering a static rather than a dynamic build.

The study has shown that linking Geant4 statically instead of dynamically allows to produce binaries running from 6% to more than 10% faster, depending on the geometry considered and the machine used. For all the analyzed geometries, e.g. GDML inner detector, GDML without EMEC and GeoModel, by performing the computations on the cluster, slightly more marked improvements have been achieved (the difference is around 3% approximately) in terms of execution time comparing the static build and the dynamic build.

This difference can be explained by considering the fact that, as can be seen from Tab. 5.1, the two machines have different features: the cluster has 128 GB of RAM compared to the 64 GB of the CERN machine. In addition, the L3 cache of the cluster's CPU is larger than the corresponding cache of the CERN machine. According to what has been discussed in Chapter 4, this ensures that a large amount of data can be stored closer to the CPU and the direct effect is a lower rate of cache misses, hence a reduced latency.

The single dynamic library approach exhibits an increase of 10% in the execution time; this degradation in performance, as previously discussed, can be ascribed to the trampoline/lookup table mechanism of dynamic linking (see Sec. 4.2.2.2). Based on the available results, this new build type, tested here for the first time, is not recommended as a viable choice for improving the execution times of the ATLAS full simulations.

5.3.2 Optimization flags

Significant differences are found when setting the optimization flags according to Sec. 5.2.2. In particular, a performance degradation of almost 23% is observed for the dynamic build type with flag `-Os`.

However, the static build type remains the best approach to achieve the largest improvements and is recommended for future studies.

It is also important to stress that the overwriting of the compiler's flags is performed silently by CMake during the configuration phase. It is therefore necessary to ensure that the correct flags have been enabled before invoking the compiler. This can be done by manually inspecting the output of the CMake command.

5.3.3 Evolution of GCC compilers

The studies presented in this work have been carried out with three different GCC compiler versions, namely 4.8.5, 6.2.0 and 8.2.0. The evolution of these three releases is articulated over the course of three years and reflects the evolution of machine hardware as well as programming language standards [148].

A complete review of the history of the GCC compilers, available from the official changelogs [149–152], is beyond the scope of this thesis. However, many of the improvements in each version target relevant areas of the Geant4 simulation code:

- From GCC 4.8.5 to GCC 6.2.0:
 - support for C++14 standard
 - improved CPU register allocation
 - new options for detecting invalid memory access and undefined behaviours
 - improvements in optimizations applied at link time (Link Time Optimization, or LTO) and to Feedback Driven Optimization (FDO)
 - better support for parallelization libraries
- From GCC 6.2.0 to GCC 8.2.0
 - support for new C++17 standard
 - improved detection of buffer overflows and invalid memory access
 - new default options in optimization flag `-O3` for loop unrolling
 - improved treatment for redundant code
 - improved verbose feedback to the user to guide the code development.

These are only a few example of the typical improvements introduced from one compiler release to another and they mostly aim at improving memory efficiency either by improving the detection and rearrangement of critical areas of the code or by improving the interaction with the user, so that the code can be refactored to achieve better compliance with the standards and to facilitate the aforementioned detection of critical paths. Based on the past evolution, it is reasonable to assume that newer versions of GCC compilers will lead to even better performance.

5.3.4 Primary particles

In order to investigate the impact that static and dynamic builds have on interactions of different complexity, several primary particles (protons and charged pions) of different energies (10, 20 and 50 GeV) have been considered.

For all the primary particles analyzed, a decrease in the simulation execution time is observed for the static build, when compared to the dynamic case.

The observed speed-up is more than 6% in the case of 50 GeV protons tested with the full ATLAS geometry and exceeds 10% in case of 20 GeV protons tested with GDML geometry.

5.3.5 Final remarks

Geant4 is the common backbone to the majority of HEP simulation infrastructures. Full simulations, performed with Geant4, are usually complemented with fast simulation options, integrated within Geant4 and provided by different frameworks.

The upgrades in view of the HL-LHC runs set ambitious computational requirements. An intensive R&D effort is ongoing to upscale the ATLAS simulation resources to meet the demanding CPU requirements. The objective is to achieve tangible simulation speed-ups and several options are under investigations:

- efficient use of parallel computing techniques;
- GPU-based computing;
- Fast simulations.

Fast or parameterized Monte Carlo simulations will be heavily exploited to boost the simulation performance. Full simulations will be improved by exploiting the opportunities offered by multi-threading, for which Geant4 is enabled since version 10.0 (at the time of writing, the latest Geant4 release is version 10.7). This could also have an impact on how the WLCG resources can be used.

In addition to the improvements on the multi-threading front, the use of heterogeneous hardware environments is also being considered. This would allow to benefit from GPU-based High Performance Computing (HPC) clusters, although Geant4 is not currently supporting this. Machine learning techniques, which can already take advantage of GPU computing, are being explored for fast detector simulations.

Attention is also focused on the details of the Geant4 full simulation workflow. Following is a list of active lines of research:

- intrinsic performance improvements coming with newer Geant4 releases, e.g. 5% speed-up from new GammaGeneralProcess and 7% between versions 10.5 and 10.6 due to logarithmic calls reduction in EM physics;
- optimization of Geant4 physics parameters, statistical decision algorithms and range cuts;
- geometry optimizations by means of new shape definitions and more efficient design;
- optimization of Geant4 run time performance by means of static linking of the Geant4 libraries.

A special mention must be reserved for the GeantV project [153]. Its main objective was to exploit modern CPU vectorized registries by refactoring the Geant4 simulation workflow.

Instead of repeatedly executing a sequence of tasks for each track, the new approach aimed to regroup data based on the required tasks. The benefit of this rearrangement is an increased locality of data in memory: information is grouped in time and by type of operation, thus opening new opportunities for parallel execution, such as those offered by SIMD and SIMT, as discussed in Sec. 4.2.1.1.

The expected speed-up factor is between 2 and 5 on modern CPUs. In addition, considering the variety of existing high-speed hardware devices, such as GPUs, FPGAs, etc., the future developments in this direction must also aim at ensuring efficient hardware portability and cross-compatibility.

Chapter 6

Optimization of detector geometry description

In this chapter, the detector descriptions produced by the `GeoModel` software are investigated. The objective is to achieve a reduction of the simulation execution time by replacing the existing geometry description of some detector components with newer shapes characterized by smaller memory and computational footprints. In particular, the attention is focused on the geometry description for the Transition Radiation Tracker (TRT). This preliminary study is part of a larger ongoing effort by the Geant4 Optimization task force [154, 155].

The study is subdivided in two steps: first, the goodness of the `GeoModel` model is assessed taking the approved ATLAS CAD drawings as a reference; then, the existing geometry is modified and the effects on execution time are quantified.

The investigations presented in this chapter were carried out using the `FullSimLight` Geant4 simulation [143]. Similarly to the `HepExpMT` benchmark simulation used in Chapter 5, the `FullSimLight` code is a general-purpose simulation, instrumented with diagnostic code to measure execution time, simple energy deposits, etc. Additionally, the `FullSimLight` simulation supports `GeoModel`-based geometry descriptions provided in the form of shared libraries (`.so` files).

6.1 Sources of geometry descriptions

As discussed in Chapter 4, Geant4 allows the definition of the detector geometry in multiple ways. Simple descriptions can be effectively provided with the Geant4 C++ classes, but for

complex objects it is convenient to import the geometry definitions produced with other tools, such as the `GeoModel` package presented in Chapter 5.

The complete set of approved ATLAS CAD drawings is available via the CERN’s Engineering Data Management Service (EDMS) and the CERN Drawing Directory (CDD) [156, 157]. These resources are of particular importance to validate the geometry description tools, as they provide official references to otherwise inaccessible detector components.

For the studies presented in this work, the relevant documents are listed in Tab. 6.1.

Table 6.1: List of relevant EDMS documents. A brief description is given together with the document file name.

	Document description	Document ID
1	Straw coordinates for module type 1	atlitb1_0004-v0.pdf
2	Straw coordinates for module type 2	atlitb2_0004-v0.pdf
3	Straw coordinates for module type 3	atlitb3_0032-v0.pdf
4	Wireframe coordinates for BSS vertices	atlitb__0028-vAB_plt_cpdf.pdf
5	Details of the BSS structure	atlitb__0024-vAD_plt_cpdf.pdf
6	Details of internal module components and clearances	atlitb1_0001_vAA.pdf

The `GeoModel` package provides a parameterized implementation of the detector geometry. This allows flexibility in case of design variations, as updating a limited set of values is sufficient to propagate the change to all the affected volumes.

The ATLAS Detector Description Database (DDD) [158] is the centralized storage location for the ATLAS geometry parameters. The evolution of each detector component is tracked and details of old geometries can be retrieved.

The Athena framework takes advantage of this database to bootstrap the `GeoModel` geometry. Alternatively, specific versions of the parameter sets can be saved to an XML file, which can then be used in standalone applications.

6.2 Assessment of `GeoModel` accuracy

To ensure agreement between the approved drawings and the simulated geometry, it is necessary to ensure that:

1. no distortions are introduced in the TRT modules;
2. their size is in agreement with the approved version.

The CAD drawing for the three module types is presented in Fig. 6.1. The corresponding part of the GeoModel1 model is shown in Fig. 6.2.

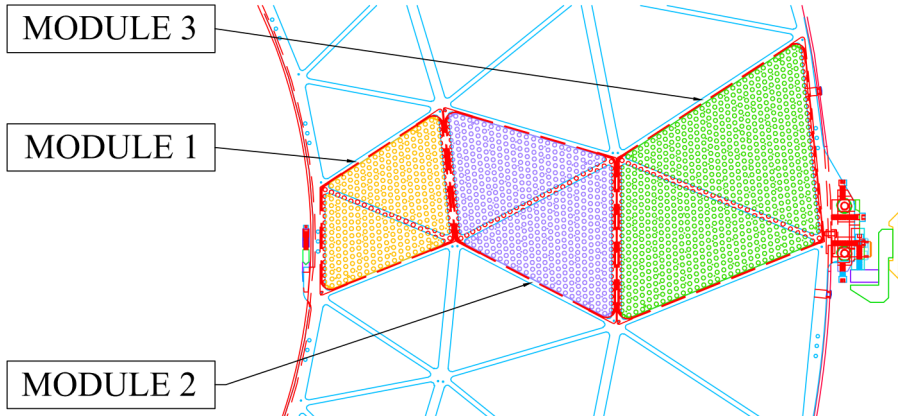


Figure 6.1: Layout of the three types of modules as mounted inside the Barrel Support System [109]. Module types 1 and 3 are oriented towards the positive y axis, whereas module type 2 is oriented downwards.

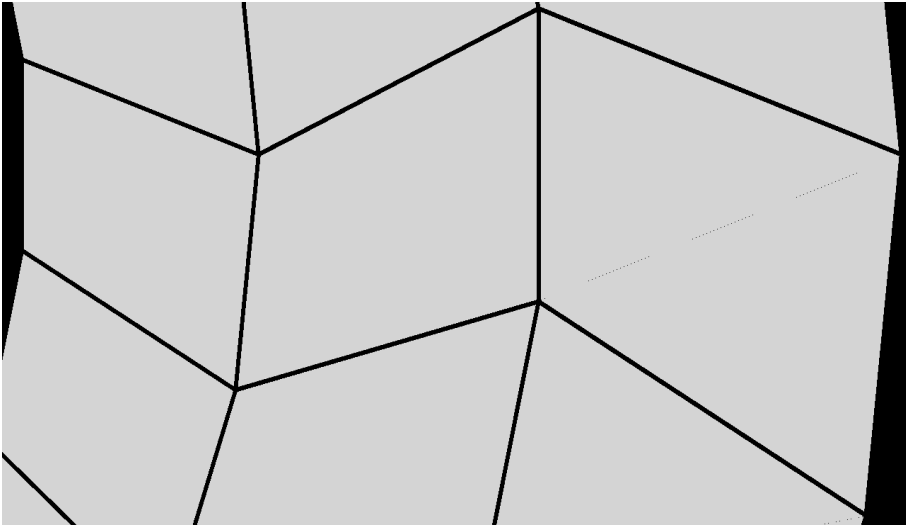


Figure 6.2: Layout of the three types of modules as represented in the GeoModel1 model. Unlike the CAD module, the module orientation is reversed, with module types 1 and 3 pointing downwards and module type 2 pointing upwards.

Despite the availability of all the (x, y) coordinates for all the TRT modules, a direct comparison with the simulated geometry is not achievable due to the different layout of the modules: in the CAD representation, module types 1 and 3 are oriented towards the posi-

tive y axis, whereas module type 2 is oriented downwards; the GeoModel model is instead reversed.

As discussed in Chapter 3 (see Fig. 3.7 on page 57), the TRT barrel is capped at the two ends with a mesh-like structure called Barrel Support System (BSS). As shown in Fig. 6.3, the CAD coordinates of the BSS allow to identify points matching orientation of the GeoModel modules. The objective is to ensure that the modules' sides are parallel to the BSS openings according to the CAD drawings; this allows to check the GeoModel description against the BSS points.

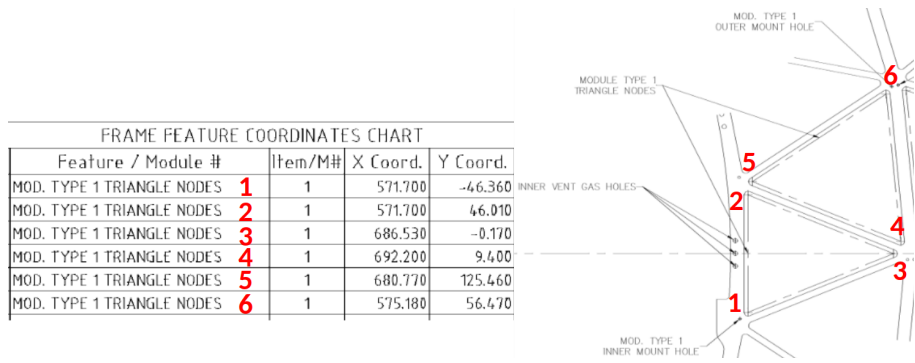


Figure 6.3: Coordinates of six vertices of the BSS mesh corresponding to the first module of type 1.

The outline of the BSS triangle shapes superimposed to the outline of the GeoModel module is shown in Fig. 6.4.

The parallelism of the TRT module sides is checked by calculating the slope of the lines connecting the vertices and is found to be exact for all the three module types.

The Gauss’s formula for the area of a generic polygon is applied to calculate the cross-sectional area of the TRT modules. Cartesian coordinates are known both for the CAD and the GeoModel representations.

The CAD modules have a larger area than their GeoModel counterparts:

- for module type 1, the difference is 3%;
- for module type 2, 2.4%;
- for module type 3, 1.95%.

This is due to a grid of 2.5 mm-wide gaps surrounding all the TRT modules which is not present in the approved drawings.

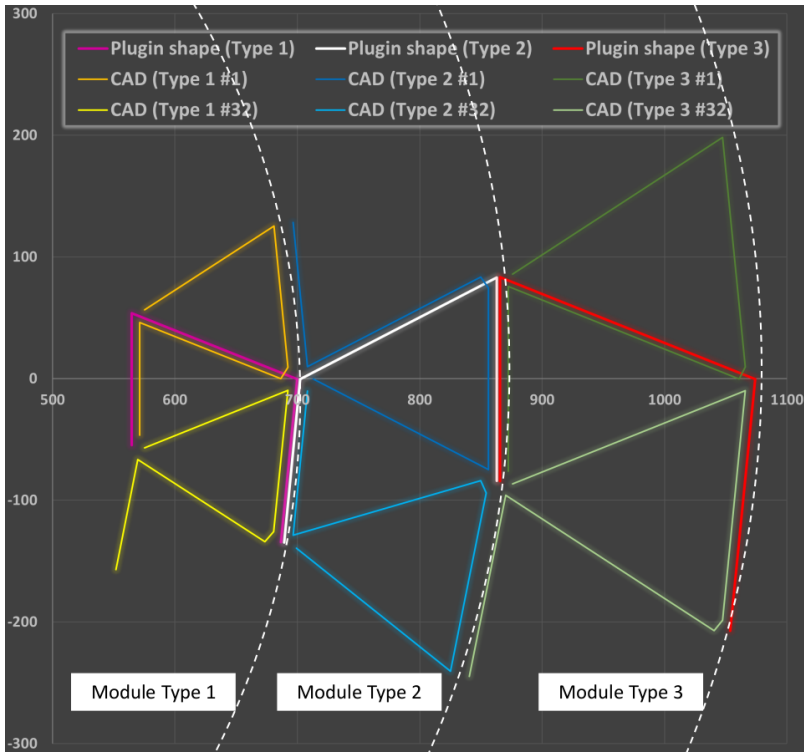


Figure 6.4: Superposition of CAD and GeoModel module outlines. The purple, white and red lines represent the shape of the GeoModel modules, whereas the yellow, blue and green lines show how to select the BSS vertices in order to assess the alignment of the two representations.

Private communications with TRT experts indicated the possibility of changes in the actual TRT layout during the tracker's commissioning phase, which have not been propagated to the CAD models. In particular, the gaps observed between the TRT modules originate from the insertion of plastic grommets to allow a flushing gas flow path [159].

It must be stressed that gaps of the order of 1 mm could also be compatible with the clearances stated in document 6 from Tab. 6.1, but a conclusive decision on the appropriate representation has not been reached. A specific support request was submitted targeting various layout imperfections [160]. However, these discrepancies are considered small enough to allow the continuation of the assessment of the performance using newly defined shapes.

6.3 Implementation of new shapes

As discussed in Chapter 3, the TRT is composed of 96 trapezoidal modules grouped in 3 types characterized by an increasingly larger cross-sectional area.

The `GeoModel` algorithm to construct the inner structure of the TRT barrel is articulated as follows:

1. one copy of each type of module outer shell is constructed by the method `makeModule` (defined in the `TRTDetectorFactory_Full` class);
2. this prototype is placed with absolute coordinates in correspondence of the first module (see Fig. 6.5). The coordinates of the four vertices are obtained from the ATLAS DDD (see Sec. 6.1);

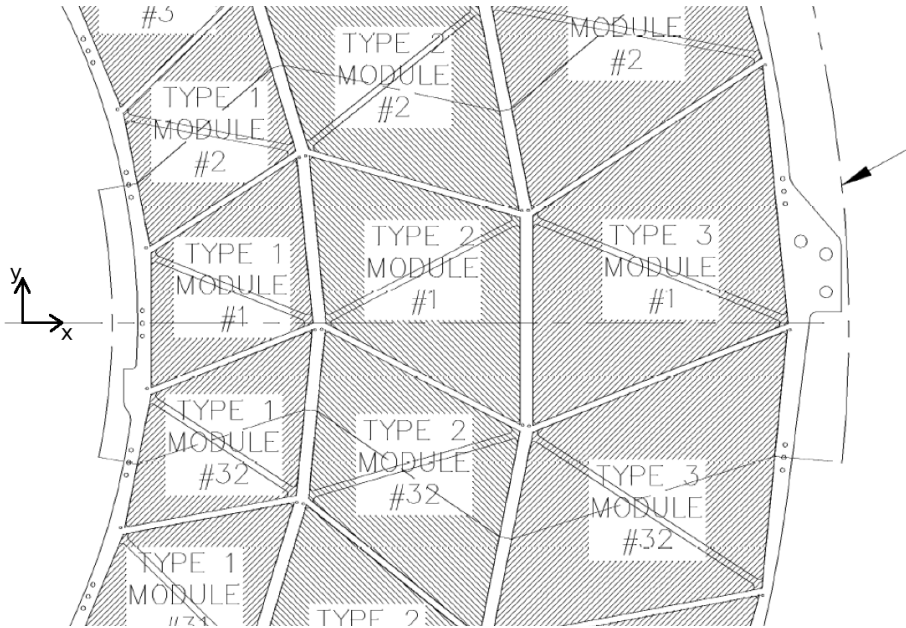


Figure 6.5: Layout of TRT modules with respect to the BSS. For each type, modules are numbered counter-clockwise, starting from the positive x axis (ATLAS reference frame) and centered around $y = 0$.

3. a transformation (rotation + translation) is defined to deploy the remaining copies to their absolute locations. Copies are not created within the `GeoModel` class; instead, they are only instantiated at runtime;

4. all the straws for a single module are defined and placed around the origin. They are laid out according to the pattern represented in Fig. 3.10. A total of 1642 straws are defined and the (x, y) coordinates are also retrieved from the Detector Description Database;
5. straws are assigned to the prototype module shell, which becomes their *mother* volume;
6. when the shells are replicated and positioned, the required copies of the straws are also created and placed automatically.

As anticipated, the method `makeModule` is responsible for the creation of the TRT module shells and is the target of the investigations presented in this section.

The current implementation takes advantage of boolean solids. Four (x, y) pairs corresponding to the four module vertices are passed to the function when invoked.

The two diagonals are calculated to identify the orientation of the module. Based on this, the trapezoid is divided in two isosceles triangles, in a fashion resembling the BSS structure (Fig. 6.4).

The prism-shaped volumes are then merged together by the common diagonal into a single solid with the desired trapezoidal cross-section. The resulting geometry is shown in Fig. 6.6 and a detailed view of the straws contained in the module shells is presented in Fig. 6.7.

In recent developments, two new shapes have been made available for `GeoModel`: the *arbitrary trapezoid* `Arb8` and the *boundary representation* `BRep`. The former requires a total of 8 points to be specified, 4 vertices belonging to the $-h/2$ plane and 4 points belonging to the $+h/2$ plane¹, whereas the latter only requires the 4 vertices describing the trapezoid cross-section to be specified.

The implementation of the two shapes is available on GitLab [161, 162].

The different module representations were tested with the `FullSimLight` benchmarking code, run with a single thread on the CERN standalone machine used for the studies in Chapter 5 (see Tab. 5.1). Each test has been repeated 10 times. As primary particles, $10^4 e^-$ at 10 GeV have been considered. The code has been compiled against Geant4 v10.6.2 with GCC 8.2.0.

The results for this comparison are summarized in Tab 6.2. A positive improvement of 1.5% is observed for the `Arb8` representation, whereas the `BRep` solid exhibits a minor degradation with respect to the reference boolean solids.

¹Here h represents the total length of the solid measured along the axis perpendicular to the trapezoidal surfaces.

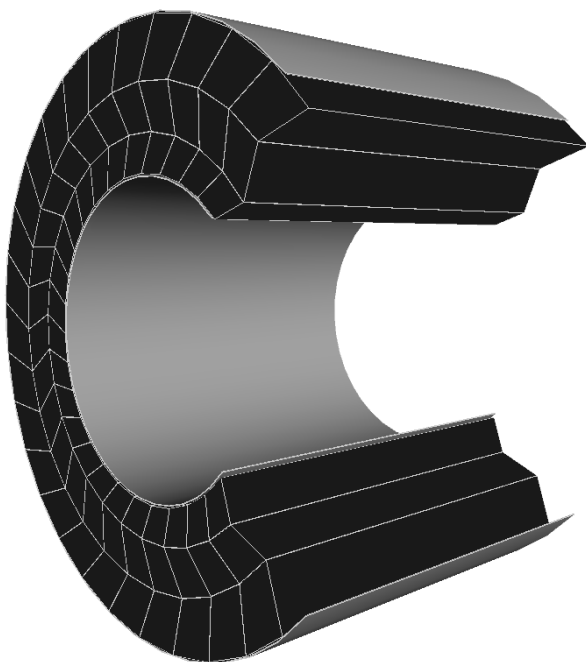


Figure 6.6: Cut-out of the 3D model of the TRT built with the GeoModel toolkit.

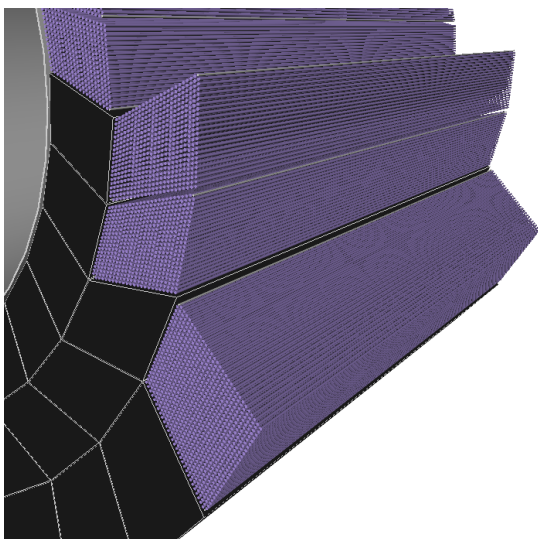


Figure 6.7: Detail of the straw tubes placed inside each module.

Table 6.2: Execution times for three types of solid representations. The current implementation with boolean solids is taken as reference. The standard deviations are of the order of 0.3%.

Module shapes	Execution time (s)	Improvement
Boolean solids	1663	Reference
Arb8	1638	1.5%
BRep	1675	-0.7%

6.4 Conclusions and outlook

The existing implementation of the ATLAS TRT has already reached a high quality standard, both in terms of geometrical accuracy and computational efficiency. The use of parameterized `GeoModel` methods and functions ensures a consistent and reproducible definition of the volumes. However, the computational improvements necessary for the foreseen workload increase require even higher efficiency to be achieved.

The results presented in this chapter have been obtained with a standalone `Geant4` simulation. It is therefore advisable to extend the investigation to the use of an Athena-based code.

Lastly, the preliminary studies carried out target only a small part of the whole ATLAS detector geometry. However, the possibility to achieve a reduction of the execution time is a positive result, clearing the way towards an extension of this approach to other detector subsystems. Currently, a similar effort is being carried out for the optimization of the EMEC.

Part III

Investigations of statistical tools for Trigger Level Analysis

Introduction to Part III

As mentioned in Chapter 2, there are several experimental evidences supporting the existence of Dark Matter (DM). A variety of astrophysical measurements ranging from the measurements of the cosmic microwave background, over the rotational speed of galaxies to gravitational lensing measurements have even allowed us to estimate its abundance: DM is five times more abundant than ordinary matter.

However, the nature of DM remains largely unknown and, in principle, DM might not be composed of particles at all. The fact that ordinary matter actually is, however, provides a valid reason to consider a particle description of DM as well. None of the particles discovered so far have characteristics that would explain their existence and abundance: DM candidates are dark, stable on the cosmological scales, non-relativistic and do not interact with any existing detector.

Searches for DM constituents are performed using different complementary approaches that rely all upon the existence of interactions between the colliding SM particles and their DM counterparts. Direct detection (DD) searches are based on DM-nucleon/electron elastic scattering interactions and they look for collisions between Galactic DM and underground targets. On the contrary, indirect detection (ID) experiments rely on the detection of SM particles produced in the annihilation or decay processes of DM in the universe.

Dark Matter could also be produced at particle colliders which had a crucial role in the long process of discovering ordinary matter. Several DM searches are ongoing at LHC. However, new interesting features are expected to be found at high energies, so triggers usually exclude lower-energy events. Moreover, due to the large event rates, full event reconstruction at all energy scales is not feasible from a computational standpoint. Thus, only a small fraction of events below these energy thresholds can be recorded, penalizing those signals with lower-energy signature.

One way to overcome the impossibility of the trigger system to record fully-reconstructed low-momentum signatures is the implementation of a novel data analysis technique, where only a fraction of the full event is saved in a dedicated data stream. The small size of

these partial events opens up the possibility of recording much higher event rates. This is particularly important because Dark Matter particles may have low masses.

In this third and last part of this thesis, the attention is focused on the di-jet resonance searches at sub-TeV masses performed by the ATLAS Trigger-object Level Analysis.

Chapter 7

Dark Matter searches at colliders using di-jets

At the origin of the Dark Matter (DM) searches at colliders is the assumption that, if the DM particles were characterized by non-zero couplings to SM particles, they could be produced in collider experiments. This generic assumption does not limit the number or type of DM-ordinary matter interaction channels.

Indeed, the DM literature ranges from fully specified, self consistent theories such as SUSY to *simplified models* [76]. In addition, many other models exist: *asymmetric DM models* (theorized following the same principle of matter/anti-matter [163]), *strongly interacting massive particles* (SIMPs) [164] and *long-lived particles models* [95], to name a few.

In Sec. 7.1, the properties of the *simplified models* are presented. They can capture features shared among many models and, for this reason, represent an excellent technical support for the design of generic searches¹. In particular, the attention will be focused on those simplified models involving an additional DM mediator, responsible for the interaction between DM particles and SM fermions.

The non-Abelian nature of QCD leads to two main features: the confinement and the asymptotic freedom. Because of confinement, free partons cannot be detected directly. Instead, the experimental signatures of quark and gluons are the so-called *jets*. A jet is a spray

¹It should be emphasized that simplified models are disadvantageous when describing collider signatures expected by more complete paradigms. To solve this problem, an intermediate group of models with increasing complexity has been considered: several *less-simplified models* have been developed in order to describe more specific features. This category includes, for example, models characterized by a flavor violation structure more complex than that of SM or coannihilation models where two species of dark matter particles with compatible masses are present in the final state [165]. Other examples of less-simplified models are described in [166–168].

of highly collimated particles and only exists as a reconstructed object. Dedicated *clustering* and *jet finding* algorithms group (or *cluster*) detected objects based on their kinematic properties, thus defining the jets.

Di-jets, final states consisting of two energetic jets, are the objective of the analysis presented in this thesis. In Sec. 7.2, the most relevant reconstruction algorithms currently used by ATLAS are concisely reviewed.

In Sec. 7.3, the general workflow of a typical ATLAS di-jet resonance search is summarized.

7.1 Dark Matter signatures at LHC

7.1.1 Theoretical framework

Based on the models and signatures currently under investigation at the LHC and, in fact, at colliders in general, it is straightforward to conclude that the dark sector and, with it, the theories, are characterized by arbitrary complexities. Some BSM theories address inherent SM problems, as in the case of supersymmetric (SUSY) theories (Chapter 2). These theories postulate partner particles to all SM degrees of freedom and introduce convenient properties, such as coupling unification, into the Standard Model.

This section deals with the so-called *simplified models*, which treat the mediator of DM-SM interaction as a new DM particle itself. Such models can be seen as a development of the most elementary portal model according to which the Standard Model is extended with a single DM particle and SM bosons such as Higgs and Z bosons act as mediators of the interactions between SM and DM particles. The simplified models, which become relevant when the collision energy is near or higher than the mediator mass [76], can therefore be considered complementary to the portal models. They describe the leading-order collider phenomenology, overcoming the limitations of some DM models dealing with the production of invisible particles in the context of Effective Field Theories (EFT) [169]².

Among the models based on the existence of the BSM mediator sector, it is possible to identify two macro-categories:

- models in which the mediator sector is composed by an extended Higgs sector plus an additional DM mediator with either spin-1 or spin-0 [95];
- models in which the mediator sector is composed of a single particle of either spin-1 or spin-0.

²EFTs provide a framework to explore a wide range of different physical processes, while remaining agnostic about specific mediator mechanisms.

A detailed review of each model is beyond the scope of this discussion; a schematic summary is available, for instance, in [95]. The DM simplified models presented in this chapter assume DM particles to be Dirac fermions and are classified considering the properties of the mediator sector. Three different types of spin-1 mediators can be distinguished:

- a baryon-number changing charged mediator [170, 171];
- a flavour-changing neutral-current mediator [165, 172];
- a neutral mediator [173–175].

In particular, the simplified models assuming a neutral mediator (Fig. 7.1) are crucial for this thesis work, as they are the objective of the Trigger Level Analysis investigations.

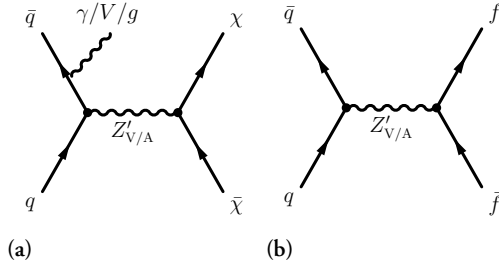


Figure 7.1: Schematic representation of the dominant production and decay modes for the V/AV model [95].

This type of models include both vector boson (Z'_V) and axial-vector boson (Z'_A) mediators and consist of a simple extension to the SM with an additional gauge symmetry $U(1)$ [95]. As shown in Fig. 7.1(a), the Z' dark mediator is produced by the annihilation of a pair of quarks or gluons and can either decay into a pair of DM particles³, or into a pair of quarks or leptons (Fig. 7.1(b)). In the first scenario, together with the two DM particles (χ), an additional visible object is produced as Initial State Radiation (ISR); the latter can be a jet, a photon or W/Z SM bosons. These two scenarios give rise to different experimental signatures of DM presence, as explained in Sec. 7.1.2.

In their most general formulation, the Z' models are characterized by five different parameters:

- the mass of the DM particle (m_χ);

³It is possible to assimilate these DM particles to WIMPs (Chapter 2).

- the Z' mass (M_{med});
- the couplings of the Z' boson to quarks (g_q);
- the couplings of the Z' boson to leptons (g_l);
- the couplings of the Z' boson to invisible particles (g_χ).

To show the complementarity between the constraints from visible and invisible searches in different mediator decay channels, three benchmark coupling scenarios have been considered so far:

- leptophobic Z'_V and Z'_A scenario: $g_q = 0.25$, $g_l = 0$, $g_\chi = 1$;
- leptophilic Z'_V scenario: $g_q = 0.1$, $g_l = 0.01$, $g_\chi = 1$;
- leptophilic Z'_A scenario: $g_q = 0.1$, $g_l = 0.1$, $g_\chi = 1$.

These values are selected considering the current LHC search sensitivity [95].

7.1.2 Experimental signatures

In an event, undetectable WIMPs DM particles, produced together with visible objects emitted as initial state radiation (Fig. 7.1(a)), lead to an imbalance in the momentum measured in the transverse plane. This quantity, called missing transverse momentum (P_T^{miss}) or missing transverse energy (E_T^{miss}) is a key property for WIMPs candidates searches at LHC [95]: due to the numerous models available, the preferred strategy of many searches is to remain as agnostic as possible about any model and to only aim to detect excesses of missing energy over the SM-only background.

DM candidates can also be produced at LHC via an s-channel exchange of DM Z' that could decay back into SM final hadronic or leptonic states⁴ (Fig. 7.1(b)). In this case, the production of DM mediator appears in the form of resonances in the di-jet (or di-lepton) invariant mass spectrum or in the di-jet angular distribution.

⁴In this context, only hadronic final states will be investigated; resonance searches aiming to constrain mediator couplings to leptons are described in [176, 177].

QCD calculations predict that di-jet production is characterized by smoothly falling invariant di-jet mass m_{jj} distribution⁵; however, new massive particles like Z' , decaying into di-jet final states, could result in additional contributions. These BSM phenomena are expected to manifest themselves via resonance peaks which appear as bumps on the steeply falling down m_{jj} spectrum⁶.

The research work presented in the next chapters is focused on resonances in the di-jet invariant mass spectrum. Searches for new physics using the di-jet mass spectrum at collider experiments have a long tradition: they were already performed at the earlier colliders back to the 1980's such as SPS and Tevatron. Also the UA1, UA2, CDF and Do collaborations have publications on this type of research [180, 181] and, currently, same kind of studies are performed at the LHC in ATLAS and CMS. Di-jet resonance searches exploit the expected absence of bumps in the di-jet invariant mass spectrum to estimate the SM background (Sec.7.3). A fit to data is performed in order to minimize the uncertainties related to the theory and to the model itself. This approach allows the observation of low-rate localized excesses (width mass ratio up to 15% in ATLAS) from resonant di-jet production [178, 179].

The large production rate of jets at the LHC allows to investigate di-jet physics on a broad mass range: in ATLAS, traditional di-jet searches are sensitive starting from masses of 1.5 TeV [92, 95, 178]. The results of these searches are expressed as upper limits on the quark-mediator coupling as a function of the mediator mass. The most stringent limit, given by ATLAS, predicts $g_q = 0.08$ for a mediator mass of 1.5 TeV [92, 178].

Pairs of low- p_T jets are produced at a very high rate and, to avoid the saturation of the trigger bandwidth, high p_T thresholds have to be set (Chapter 8). This causes a loss of sensitivity of traditional di-jet searches to low-mass mediators.

To overcome this limitation, it is possible to either:

⁵Starting from an event with a pair of jets and selecting the *leading* and *subleading* jets, i.e. the two jets with the highest p_T in the event, it is possible to obtain a *di-jet* by performing the four-vector addition of the two jets' four-vectors. The di-jet mass is then obtained according to Eq. 7.1:

$$m_{jj}^2 = E_{jj}^2 - p_{jj}^2 \quad (7.1)$$

where the jj indices indicate that all the variables of the Eq. 7.1 are di-jet four-vector quantities. The invariant mass is a crucial experimental observable for many researches since it is independent from the inertial frame and, in case of a particle decay into two jets, it is a conserved quantity.

⁶Bump search strategy turns out to be limiting in the presence of broad resonances since it would be difficult to distinguish these resonances over the smoothly falling QCD background. To address these limitations, both CMS [178] and ATLAS [179] collaborations have developed a new approach by defining a new variable, the jet angular distribution χ , which acts as a discriminant from DM signal and QCD background. χ is not sensitive to the resonance width but has been defined in such a way that QCD events have an angular distribution independent of χ while it is expected that DM mediators would preferentially produce events characterized by low value of this variable.

- apply an event selection in which the two jets are highly boosted and merged together due to recoil against an additional hard ISR jet. In this case, it is possible to obtain, in principle, events with three high- p_T jets that are able to pass the standard trigger thresholds, even if very light. Thanks to this technique, adopted by CMS [178] and ATLAS [182], the di-jet invariant mass between 50 GeV and 300 GeV has been covered [183, 184];
- implement a novel approach where the limitations of the trigger system are overcome by reconstructing physics objects used for the analysis (e.g. electrons, photons, jets ...) within the trigger system itself. Only a fraction of the full event is saved in a dedicated trigger data stream; this allows to distinguish between these *trigger objects* and the *offline objects*, which instead are reconstructed using the full detector information and the final calibration constant [185]. This technique has been adopted by several LHC experiments: it is called *Turbo stream* by the LHCb experiment [186], *Data Scouting* by the CMS experiment [187] and *Trigger-object Level Analysis* (also known as Trigger Level Analysis, TLA) by the ATLAS collaboration.

As in the case of searches for invisible final states, also analyses looking for visible final states found no significant deviation from the expected SM background; therefore more thorough studies are needed.

According to the recommendations of the *ATLAS/CMS Dark Matter Forum* [188], scans over the mediator and dark matter particles masses are performed; this allows to provide results in terms of di-jet search contours for 95% upper limits on the coupling g_q as a function of the resonance mass $m_{Z'_A}$ (Fig. 7.2).

Limits from the E_T^{miss} analyses have been derived using a mixture of simulated signal samples; instead, resonance searches have derived the model-independent exclusion limits in the mediator mass plane from the limits placed on hypothetical signals producing a Gaussian-shape contribution (characterized by different widths) to the m_{jj} distribution.

7.2 Jets reconstruction

Most of the inelastic p-p collisions occurring at the LHC result in the production of quarks and gluons which, as explained in Chapter 1, cannot exist in isolation since they are subjects to color confinement. Jets are collimated sprays of hadrons, which arise from the showering and hadronization of quarks and gluons. These particles, propagating through the ATLAS detector, leave tracks and energy which, by means of dedicated algorithms (Sec. 7.2.1 and Sec. 7.2.2), can be combined to form the reconstructed jets.

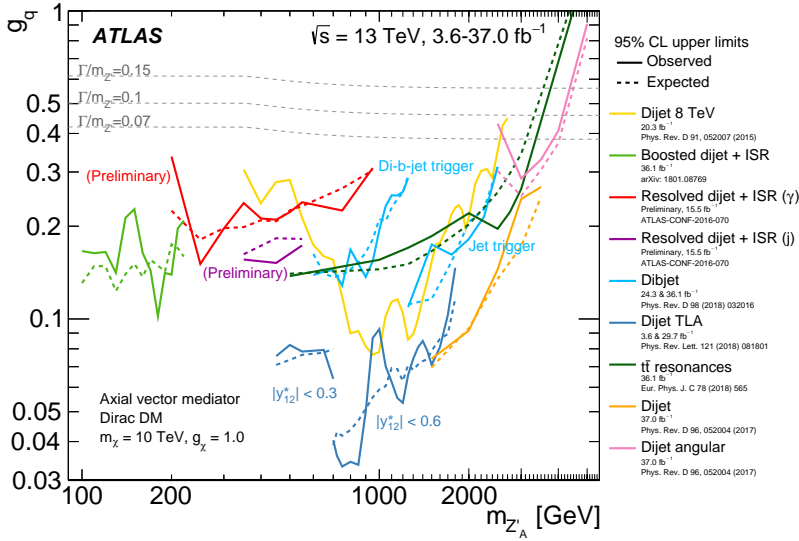


Figure 7.2: Exclusion contours found by different ATLAS searches for a DM simplified model that considers the leptophobic axial-vector Z'_A . Dotted lines indicate the expected limits from each search; coupling values above the solid lines are excluded. Two different datasets with different selections in the rapidity have been employed by the TLA; this explains why TLA limits are splitted into two parts [95].

In an experimental context, however, it is possible to observe the detector measured jets and the final purpose is to relate the latter with the first to probe the initial partons.

There are a variety of different ways to define and reconstruct jets depending on the detector design and the physics of interest [189]. However, all these approaches tend to be structured in two main phases:

- identification of regions or objects of interest;
- build jets using the objects selected in the previous step.

Once a set of jets is obtained, in order to provide an accurate description of the event, it is also necessary to perform a rather complex calibration process (Sec. 7.2.3).

7.2.1 Input to jet algorithms

ATLAS mainly uses topologically-grouped noise-suppressed clusters of calorimeter cells, called *topo-cluster*, as input to jet reconstruction algorithms. Other approaches exist, con-

sidering, for instance, the charged particle traces in the inner detector, or a combination of data from inner detectors and calorimeters [190]. This is the case with *particle flow jets*, which use this information to exploit the superior performance of the trackers at low energies and the calorimeters at high energies.

The ATLAS calorimeters are made up of several thousands of cells where energy is deposited. To obtain the input objects of the jet finding algorithm, e.g. the *detector objects*, the cells characterized by a signal above a certain threshold are topologically connected forming the *topo-clusters* [191].

This clustering process takes place considering the ratio of the measured energy in a single cell E_{cell} to the average expected noise in that cell σ_{noise} . In this way, it becomes possible to localize the high energy density detector regions suppressing the background of electronic noise and other sources of fluctuations such as pile-up.

Topo-clusters are formed applying a growing-volume procedure: starting from a cell characterized by a signal-to-noise ratio above a certain threshold, increasingly stringent criteria are applied to include neighbouring cells with lower energies (Fig. 7.3).

The topo-cluster formation is structured in three main steps:

- cells with $E_{cell} > 4\sigma_{noise}$, the seed cells, are selected; each of these cells forms a proto-cluster. In Fig. 7.3, these type of cells are marked with red color;
- all neighboring cells characterized by $E_{cell} > 2\sigma_{noise}$, marked in orange, are added to the proto-cluster;
- all neighbouring cells with $E_{cell} > 0\sigma_{noise}$ are added to the cluster previously identified. The latter are indicated in yellow.

Thus, topo-cluster formation is a sequence of seed and collect steps, which are repeated until all topologically connected cells pass the criteria listed above.

After having clustered the calorimeter cells into groups, a first set of calibration processes are applied to better reflect the true energy deposited by the particles. Corrections that aim to account for the effects of dead cells and noise thresholds are also considered [192]. These corrections are made through comparisons with simulated data where the detector description includes also dead material.

There are two schemes commonly used in ATLAS to perform this procedure:

- Electromagnetic calibration; this approach uses only the energy released through purely electromagnetic interactions. Since the response of non-compensated calori-

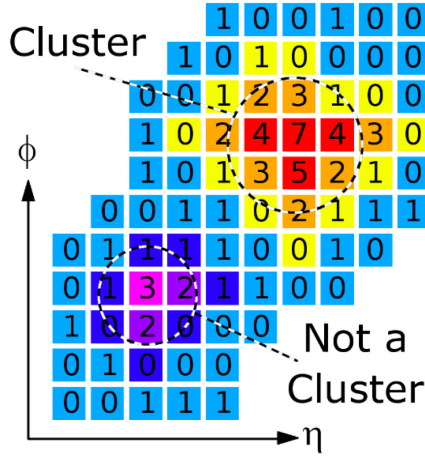


Figure 7.3: Topo-cluster algorithm in calorimeter cells. Proto-cluster cells are marked in red; orange represents cells added in step 2 and yellow cells added in step 3 [191].

meters is lower for hadronic showers, the response of hadronic deposits is considerably worse with respect to the case of electromagnetically interacting objects.

- The *Local-Cell Weighing* method (LCW) [192]; this second method compensates for the lower response of the hadronic interactions since it is able to establish the electromagnetic or hadronic nature of the shower by exploiting the distinctive features of the two types of interactions. EM showers generally have a more compact evolution, earlier starting points and smaller energy fluctuations than hadronic showers.

The four-vector containing the direction and the energy of the various topo-clusters which are part of an event must be provided to the jet finding algorithm. The topo-cluster direction is found by connecting with a straight line the center of the detector to a signal-weighted barycenter of the cluster cells while the total energy is calculated performing a vector sum of the cell energies included in a cluster.

7.2.2 Jet finding algorithms

As mentioned in the previous section, there are several methods for reconstructing jets based on different detector inputs. In this section the focus will be on jets whose constituents are the calorimeter topo-clusters.

In QCD the cross-section divergences that emerge from infinitely soft radiation and splitting at vanishing opening angles are regularized in such a way that the final results are not

influenced by these effects.

Similarly, to be experimentally useful and with a strong theoretical foundation, a jet definition algorithm should be both *infrared* (IR) and *collinear-safe*⁷.

Historically, the first jet finding algorithms had been developed by exploiting the cone-shape that characterizes the development of jets; these algorithms, after having established a cone size, added all the momenta inside it. However, they turned out to be collinear and/or IR unsafe.

Currently, *sequential jet-finding algorithms* are commonly used in ATLAS [193].

Once clusters have been identified, jet finding algorithms are applied. Their generalized formulation is based on the distance d_{ij} between clusters i, j (Eq. 7.2) and on the distance d_{iB} (Eq. 7.3) between cluster i and the beam axis B [194].

$$d_{ij} = \min \left(p_{T_i}^{2p}, p_{T_j}^{2p} \right) \frac{\Delta R_{ij}^2}{R^2} \quad (7.2)$$

$$d_{iB} = p_{T_i}^{2p} \quad (7.3)$$

where R is the typical radius of the resulting jets, $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and p_{T_i} , y_i and ϕ_i represent the transverse momentum, the rapidity and the angle of the i -th cluster, respectively.

The generalized workflow of the reconstruction algorithm is as follows:

1. d_{ij} and d_{iB} are computed for all clusters and cluster pairs;
2. the minimum of these quantities is found;
3. if the minimum distance is from the set of d_{ij} , then the four-vectors identifying the two cluster are combined into a new *proto-jet*;
4. otherwise, if the minimum is from the set of d_{iB} , then the four-vector of cluster i becomes a jet itself and is no longer considered by the algorithm;
5. the algorithm stops when a cluster j such that $d_{ij} < d_{iB}$ is no longer found.

⁷IR-unsafe and collinear-unsafe algorithms are prone to errors in case a soft gluon is emitted or a quark undergoes a small-angle splitting, respectively.

In Eq. 7.2 and Eq. 7.3, transverse momenta are raised to the power of $2p$. The coefficient p can be positive (k_T , p_T enters as a numerator), negative (anti- k_T , p_T enters as a denominator; $p = -1$ is the typical choice of ATLAS) or null (p_T is not considered; this is known as the Cambridge-Aachen algorithm). The choice of p has an effect on the sorting of small and large p_T , which in turn determines how random/irregular is the reconstructed shape of a given jet [193, 195, 196].

7.2.3 Calibration

After being reconstructed, jets⁸ need to be calibrated through a multi-step procedure. In Fig. 8.3 the steps required for calibrating small-R jets is outlined [197, 198].

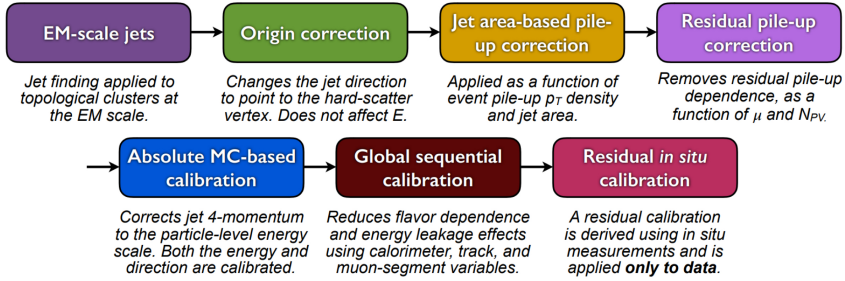


Figure 7.4: Overview of the calibration chain for small-R jets [199].

7.2.3.1 Origin correction

The small-R jets calibration process starts with the *origin correction* step. At the EM scale, the geometrical centre of the detector is considered as the origin of the jet four-momentum. The jet axis is corrected so that it points to the hard-scatter vertex of interest.

7.2.3.2 Pile-up corrections

In order to perform this step, two jets parameters are used: the *jet area* A_{jet} and the event-by-event energy density ρ . A_{jet} is evaluated for each individual jet in each event and provides a measure of the local size of the jet. ρ is a measure of the global pile-up energy of each given

⁸Jets representing quarks and gluons are called small-R jets and they are reconstructed setting $R = 0.4$. On the contrary, jets representing massive particles that decay hadronically are called large-R jets since they are reconstructed with $R = 1.0$. The analysis described in this thesis uses small-R jets [189].

event and it is subtracted from the initial jet four-vector in order to remove the pile-up energy contribution.

The pile-up sensitivity varies across the detector. Following the area-based pile-up subtraction it is therefore necessary to apply a residual correction based on N_{PV} and $\langle\mu\rangle$.

The fully pile-up corrected p_T is then:

$$p_T^{corr} = p_T - \rho A_{jet} - \alpha (N_{PV} - 1) - \beta \langle\mu\rangle \quad (7.4)$$

where μ is the average number of interactions per bunch crossing, N_{PV} represents the number of reconstructed primary vertices in the event and α and β are the derived average correction factors.

7.2.3.3 Absolute MC-based calibration

After the removal of the pile-up contribution, the jet energy is calibrated using Monte Carlo (MC) truth information. This step allows to correct for the fraction of energy from hadronic showers remaining invisible to the detector. Reconstructed jets from measured data are compared to jets built from simulated particles (*truth jets*). The Jet Energy Scale (JES) measurement depends on this correction.

7.2.3.4 Global Sequential Calibration (GSC)

Small-R jets further refine the JES applying the *Global Sequential Calibration* (GSC). In general, jet reconstruction is sensitive to the jets' components and energy distribution. The energy response of a jet initiated by a quark can differ even up to 8% compared to the response of a gluon jet. This discrepancy is corrected through several steps based on tracking, muon spectrometer information and calorimeter response; all these corrections are applied independently and sequentially in the Global Sequential Calibration step.

7.2.3.5 Residual *in situ* calibration

Except for pile-up corrections, the other calibration steps discussed so far are based on MC results. For this reason, it is extremely important to understand how well MC describes collision events, hadronization processes and particle interactions with the detector. The *in-situ* calibration accounts for the disagreement in the jet response and it has been refined over many iterations.

The first step is the di-jet η -calibration and it aims to equalize p_T of the more forward jets to the p_T of the well-measured central region jets. The η -calibration is followed by a series of p_T balance methods where objects that are well understood are used as a reference to correct the absolute scale of jets between data and MC.

The ultimate goal of detecting, measuring and calibrating the detector-object jets is to probe the initial partons. For this reason, once jets have been calibrated it is necessary to continue with a further step called *jet tagging* in order to evaluate the angular distribution of energy within the jet. This allows us to identify the origin of the jet and separate the boosted massive hadronic particle decays from QCD di-jet production [189].

7.3 Di-jet analyses workflow

Each specific analysis defines some selection criteria in order to properly construct the m_{jj} spectrum and the events that do not meet these requirements need to be rejected.

In general, the standard method of ensuring good data quality for all events is to consider the *good run lists* (GRL): a series of lists created by the ATLAS Data Quality group. There are also many other selections and checks that can be applied in the data selection and preparation phase. For example events with data integrity issues are rejected and also events with calorimeter integrity issues or tripped modules are vetoed. Then there is a whole series of events that is rejected by imposing conditions on m_{jj} , p_T and $|y^*|^9$ of leading and sub-leading jets. These conditions vary from analysis to analysis.

After all selections are applied, the resulting invariant mass spectrum must be analyzed to identify any statistically significant indication of BSM phenomena in the collected data.

The di-jet invariant mass spectrum obtained from data should be compared to a known reference, the *null hypothesis* (or *background-only hypothesis*). For the di-jet analyses, it represents the shape the m_{jj} distribution would be expected to have if only SM processes were present.

The comparison between expected and observed distributions is performed during the *search phase* (Sec. 7.3.2) and, depending on the outcome, i.e. whether or not a localized excess of events is found, the claim for a new physics discovery must be verified, or a limit on the signal strength must be set.

Uncontaminated spectra may be produced by Monte Carlo (MC) simulations, which must guarantee the same statistical errors as measured data. The current event rates of the LHC

⁹ $|y^*|$ represents the rapidity of the two partons in the centre-of-mass frame. It is defined as $|y^*| = \frac{y_1 - y_2}{2}$ where y_1 and y_2 are the outgoing parton rapidities.

are however too large to allow the production of MC-generated data with the required statistical power ¹⁰.

To bypass this limitation, *data-driven background estimation* is commonly adopted. This technique relies on the assumption that, provided that the sought resonances are narrow, measured data are mostly characterized by clean SM background even in case of excesses. It is then necessary to ensure that the background fit to the measured spectrum is insensitive to such narrow excesses.

Both the estimation of the degree of discrepancy between observed and expected data and the limit setting phase are performed using a set of pre-defined statistical tools, which can in general be classified as *frequentist* or *Bayesian*, depending on the underlying definition of probability.

7.3.1 Probability definition: frequentist vs Bayesian

Probability can be defined in two ways: the *classical* or *frequentist* method [200] and the Bayesian method and both definitions are commonly adopted for the ATLAS statistical analyses. The frequentist definition of probability is based on the ratio of the frequency with which a given outcome occurs to a large number of repeated experiments (Eq. 7.5). This definition can only be used when a large number of trials can be performed.

$$P(A) = \lim_{n \rightarrow \infty} \frac{\text{times outcome is in } A}{n} \quad (7.5)$$

The Bayesian definition of probability is based ultimately on the definition of conditional probability and Bayes' theorem takes the form of Eq. 7.6.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (7.6)$$

The fundamental property of this theorem is the dependency on the *prior* probability $P(A)$, also known as the *degree of belief* that A is true. $P(B)$ is a restriction imposed on the sample space and can be regarded as a scaling factor. Bayes' theorem can therefore be rewritten as in Eq. 7.7.

$$P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory})P(\text{theory}) \quad (7.7)$$

For a given experiment, guidelines are available to estimate the prior probability of a certain theory. Once the initial value for $P(\text{theory})$ is obtained, it can be updated with the

¹⁰This issue is the objective of Part II of this thesis.

knowledge gained from the measured data by multiplying it by the probability of having the measured data if the theory is assumed to be true. This factor, $P(\text{data}|\text{theory})$, is called *likelihood*. While, on the one hand, the Bayesian approach allows to separate a prior degree of belief from the part only depending on data, on the other hand the frequentist method does not rely on the accuracy of the prior degree of belief and is, in fact, inherently agnostic about the theory.

7.3.2 Search Phase

The objective of the search phase is to estimate the agreement of a given hypothesis with the measured di-jet invariant mass spectrum and, ultimately, to test the validity of the Standard Model. It is important to stress that the search phase is inherently frequentist. If the description of the physics processes was completely covered by the background-only hypothesis, then the observed data would only be random (Poisson) fluctuations around some nominal values. The agreement between the observed data and the null hypothesis is a measure of the validity of the H_0 hypothesis itself. This idea is quantified by means of a *test statistic* T expressing the compatibility between x (i.e. the observed data) and H_0 . The test statistic T is inversely proportional to the compatibility between the two datasets. It is a scalar quantity, thus a single value is representative of the agreement at all mass scales (the overall spectrum).

The three test statistics more frequently adopted during the search phase are:

- the Pearson's χ^2 test, which quantifies the agreement by defining the sum in quadrature of the differences between observed and expected data, normalized by the expectation [201];
- the Log Likelihood, which quantifies the conditional probability of describing a given dataset with a set of parameters θ . This test statistic, though, shall not be interpreted as a direct measure of probability, as it is not normalized to 1 [202];
- the `BumpHunter` algorithm [203], which estimates the agreement based on the analysis of the overall behaviour of neighbouring bins. Indeed, in resonance analyses, the excess in adjacent bins may only be significant if not randomly distributed. The two previously mentioned tests are not sensitive to the excess distribution in a given bin window [203]. It is important to stress that comparisons between the datasets only take into account excesses and not deficits, as the latter do not reflect the behaviour of any BSM model considered by di-jet analyses.

Regardless of the test statistic chosen, a p -value quantifying the agreement between ob-

servations and predictions must be calculated¹¹. Being the search phase an intrinsically frequentist process, *pseudoexperiments* are carried out to determine how many times a randomly generated spectrum is compatible with the H_0 hypothesis. Pseudoexperiments are performed by letting each bin fluctuate with a Poisson distribution around the background-only nominal bin count. For each pseudoexperiment, the value of the chosen test statistic is calculated. The required p -value is then the fraction of pseudoexperiments resulting in a test statistic larger than that obtained for the null hypothesis. As an additional clarification, a small p -value corresponds to a small number of pseudoexperiments giving a test statistic larger than that of the null hypothesis. Therefore, if data with a test statistic are at least as large as the largest obtained from the pseudoexperiments are observed, then they are very unlikely to originate from random fluctuations. This can ultimately lead to the rejection of the H_0 hypothesis. The p -value commonly adopted as the rejection cutoff is 0.01.

7.3.3 The background estimate

In events produced by only SM processes, the two highest-energy (or highest p_T) jets are, in general, not coming from the decay of a single particle at a fixed mass but will be the combined result of a random QCD processes. On the contrary, in the case of new physics, the leading jets in an event, come from a single source giving rise to resonance phenomena. This leads to the formation of a smoothly falling distribution for the SM background in which resonances emerge, if BSM phenomena occur.

The identification of mass excesses is dependent on the correct description of the SM-only background, whose exact form, however, is not known. It is nevertheless possible to provide guidelines for the definition of the fit function:

- it must be a monotonically decreasing function with an approximate exponential form;
- it must depend on a number of parameters matching the actual number of degrees of freedom of the described physics;
- it must not accommodate resonance-like structures in the spectrum, but it must be flexible enough to provide a good description of the di-jet spectrum at all mass scales.

Based on these considerations, the function chosen to parameterize the background is given in Eq. 7.8.

$$f(x) = p_0 (1 - x)^{p_1} x^{p_2 + p_3 \ln x} \quad (7.8)$$

¹¹It must be noted that the statistical tests adopted to estimate the agreement between observed and expected data only produce a single value representative of the entire spectrum. It is also advisable to define quantities capable of measuring the agreement on a per-bin basis. As an example, the *significance* is defined as the ratio between the discrepancy and the total uncertainty in a given bin.

where $x = m_{jj}/\sqrt{s}$ is the ratio of the di-jet invariant mass to the center of mass energy and p_i are the fit parameters. New searches are characterized by increased luminosities and extended mass ranges. The former, resulting in smaller statistical errors, allow finer features of the QCD spectrum to appear, whereas the latter introduce new constraints at previously unconstrained mass scales. Functions of the form of Eq. 7.8 are not always suitable to account for the added constraints and new parameters (i.e. more degrees of freedom) are usually introduced.

It is important to stress that the increased mass ranges represent a challenge for the traditional fitting procedures (*global fit*), i.e. those dealing with the entire range with a simple, smooth function (e.g. Eq. 7.8). Indeed, the statistical errors over a broad mass range add constraints to the fit and prevent the function from reliably describing the full spectrum.

Instead, the *sliding window fit* (SWiFt) approach is adopted, as it allows to obtain a flexible, yet insensitive to local excesses, fit to data. The method is based on the definition of a fixed-width window, which is then slid across the spectrum. At each window position, a fit is performed. Lastly, the background estimate for the i -th bin is calculated by evaluating the fit with the window centered on that bin. The effectiveness of the SWiFt method depends on the window width: a wide window results in a rigid fit, whereas a narrow window allows localized excesses (i.e. the searched signals) to be absorbed by the fit itself.

7.3.4 Validating the background-only hypothesis

Since a background-only spectrum can not be extracted from the measured data, it is necessary to estimate how well the chosen fit function represents the spectrum. This can be accomplished in two ways:

- by applying the background estimation algorithm to a reduced (*blinded*) dataset (1/4 of the events to double the error bars). This approach relies on the reduced statistical power of the blinded dataset to hide any potential new physics. The data on which the fit function is tested can therefore be treated as if they didn't contain any signal. This ensures that the fit function is not capable of accommodating localized excesses, which can then be identified when the fit is performed on the unblinded spectrum (i.e. with full statistical power).
- by producing simulated data.

These two validation approaches allow to test several fit functions, all characterized by a smooth functional form¹². For all the tested functions, a p -value is derived from different

¹²TLA-specific fit functions will be listed in Chapter 8.

test statistics (e.g. Pearson's χ^2 , Log Likelihood, BumpHunter, etc.). A given function is then discarded if the corresponding p -value calculated on SM-only data is too small. Another procedure to validate the background-only function is known as k -fold cross-validation and is applied to the full dataset. This technique allows to identify functions with more degrees of freedom than the underlying distribution, which would adapt to statistical fluctuations instead of actual physics features [204].

7.3.5 Uncertainties on the fit

Several uncertainties affect the error estimation of the fit to data. Such uncertainties can be broadly divided into those affecting the background estimation and those affecting the signal presence. The attention is focused only on the uncertainties considered by the TLA analysis (Chapter 8). Among the background uncertainties, two have the largest impact for the di-jet analysis purposes:

- uncertainty on fit parameters: when a fit to data is performed, along the nominal parameter values, uncertainties are calculated. To derive such errors it is possible either to iterate the fit procedure on a large set of pseudoexperiments to reach an empirical estimate of the fit uncertainty, or to calculate the error band analytically starting from a well-known *correlation matrix*;
- uncertainty on fit function choice: a set of validated functions may represent the same dataset with large variations. Since these differences may be larger than the statistical uncertainties associated to each function, it is necessary to take into account a new source of error. There is no standard procedure to estimate the uncertainty introduced by the choice of fit functions; the typical approach of di-jet analyses is to select a *nominal* and an *alternate* function, providing a reference and a secondary background estimate, respectively.

Similarly to the background-only case, several uncertainties on the signal exist (e.g. luminosity, beam energy, jet energy scale, jet energy resolution, etc.). The TLA analysis has included two of these [205], namely the luminosity and jet energy scale contributions, which both affect the relative normalization of the dataset and a given signal MC sample during the process of setting limits. In ATLAS, the luminosity is periodically calculated and tested and the corresponding uncertainty can be calculated from such individual tests [206, 207].

Unlike the luminosity, which is a scalar normalization coefficient, the JES uncertainty is a complex object, consisting of a total of 54 components. Not only treating all these degrees of freedom independently would be too resource-heavy from a computational standpoint, but it could lead to unphysical interpretations of the observed data. A set of orthogonal

contributions is instead calculated and their impact is computed for all signals and mass points. The JES uncertainty is applied to the nominal dataset as a $\pm 1\sigma$ shift on the m_{jj} axis.

7.3.6 Limit setting

The analysis of di-jet invariant mass spectra can essentially lead to two different outcomes: if, during the search phase, a statistically significant excess taking the form of the predicted resonance is found, then the null hypothesis is rejected and the claim of potential new physics is investigated further; instead, if no excess is found and, therefore, the null hypothesis cannot be rejected, it is possible to proceed with the *limit setting* phase. The objective is twofold:

1. finding a maximum signal strength at a given mass m_{jj} that could be hidden below the minimum threshold for statistical significance. The limit on signal strength can be translated into a limit on cross-section for a given process;
2. if the model under consideration provides a nominal value for the process' cross-section, then the intersection between the nominal cross-section curve and the upper limit curve provides the highest possible mass above which that particular interaction can be excluded.

The limit setting phase can also be defined according to the two definitions of probability [208]:

- the frequentist approach, based on comparison of multiple outcomes, relies on the values assumed by some test statistic t . Larger discrepancies from the null hypothesis result in larger values of t . Once a *detection* is excluded, the frequentist upper limit statement becomes: $P(t \geq t_0 | A > A^*) \geq 90\%$. A^* is the 90% confidence upper limit and the equation states that, assuming a signal amplitude larger than A^* , the corresponding test statistic t would be larger than t_0 with at least 90% probability;
- the Bayesian approach, being able to associate probabilities to individual outcomes, provides a conditional probability according to which, given the observed data and the assumption of the presence of a signal, the amplitude is smaller than the threshold A^* with 90% probability: $P(A \leq A^* | y, H_1) = 90\%$, where H_1 is the *signal* hypothesis.

Chapter 8

Trigger Level Analysis

Dark Matter searches at colliders have explored, and are still exploring, mass ranges extending from sub-TeV to multi-TeV energies. Regardless of the energy scale probed by the experiments, the typical DM cross-sections are extremely small.

Investigations at high energy could be particularly favourable for DM detections thanks to the low SM backgrounds. Instead, for the low-energy end of the spectrum, and in particular for the ≈ 1 TeV range explored at the LHC, it is necessary to achieve the highest possible statistics.

The typical event rates of the LHC, being many orders of magnitude larger than previous experiments, are suitable for DM searches at low masses. However, the requirements of traditional physics measurements introduce an artificial limitation on low- p_T data taking by means of the so-called *pre-scaling*. In particular, single-jet triggers with $p_T < 380$ GeV are strongly suppressed.

To recover the otherwise lost statistical power, the Data Scouting (DS) strategy has been developed.

Jet reconstruction can be performed *online* or *offline*. In both cases, topo-clusters are used as input and the two techniques differ in the type and amount of stored information. Online jet reconstruction takes place in the HLT and is only based on detector readout; offline reconstruction, instead, also requires full calibration, noise suppression and data quality flags to be taken into account. Therefore, the fewer requirements imposed for online reconstruction would allow to achieve much higher event rates for all-jet searches. The Data Scouting strategy takes advantage of the increased event rate associated to the reduced event size to improve statistics in low- p_T regions; despite the reduced amount of information saved for each event, the usability for Trigger-Level physics is not compromised.

The Data Scouting technique, firstly explored by CMS [209], has later been implemented also by other collaborations: at LHCb is known as Turbo Stream, whereas at ATLAS it is called Trigger-object Level Analysis (TLA).

This chapter provides an overview of the TLA approach applied to the search of new phenomena in the two-jet invariant mass spectrum in the region below 1 TeV. Based on [185, 205], the motivations, the challenges and the most relevant achievements of this analysis are reviewed. Lastly, studies of some performance aspects of different statistical tools applied to background fit procedures and limit setting are presented.

8.0.1 Production of TLA data at triggers

Although beneficial to measurements, the ATLAS event rate is too large to allow continuous readout, reconstruction and storage of each events. The purpose of the ATLAS Trigger and Data Acquisition system (TDAQ), represented in Fig. 8.1, is to reduce the event rate in order to cope with the limitations imposed by the data processing and transmission hardware. The selection process comprises two stages: the Level-1 (L1) trigger and the High Level Trigger (HLT).

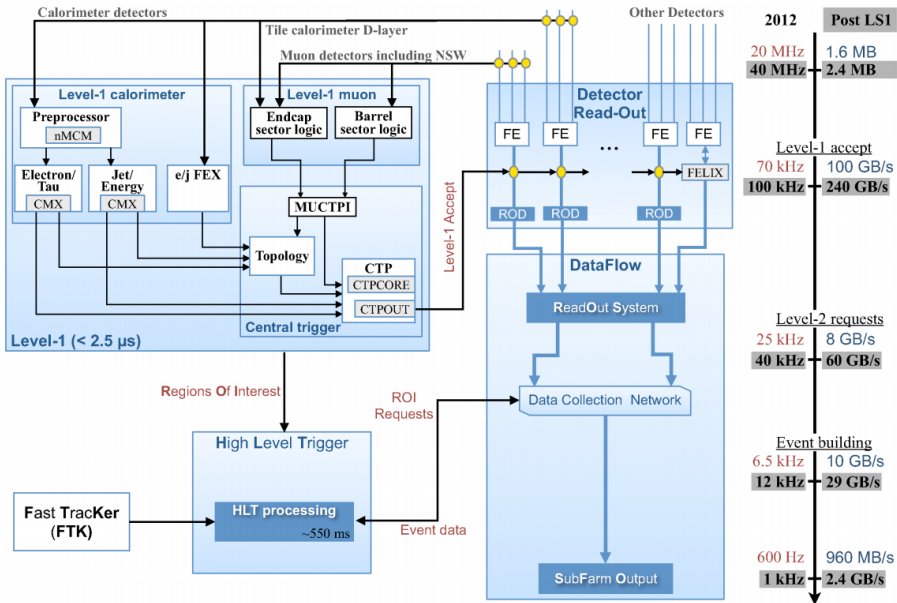


Figure 8.1: ATLAS Trigger and Data Acquisition system [210].

The L1 trigger is hardware-based and works on a subset of data coming from the calorimeter and muon detectors. Only events selected at this stage are eligible for further processing.

The latency of the L1 trigger is ≈ 2.5 s.

The input to the L1 Calorimeter trigger (L1Calo) originates as analog signals in the calorimeters, which are then fed to the Cluster Processor (CP) and to the Jet/Energy-sum Processor (JEP). The former, based on the energy deposit in the EM calorimeter, is responsible for identifying electrons, photons and τ -leptons above a given threshold; the latter, in addition to the calculation of the total and missing transverse energy, is responsible for the identification of jet candidates.

The L1 trigger selection is based on:

- event-level quantities, such as the total energy in the calorimeter;
- the multiplicity of objects above a certain threshold (e.g. the transverse muon momentum);
- topological requirements, such as the invariant mass or angular distances.

In addition, Regions of Interest (ROIs) on the $\eta - \phi$ plane are defined at the L1 level. These ROIs are further investigated at the HLT level.

The High Level Trigger is software-based and relies on a set of algorithms with average latencies of the order of 10^2 s. Data is first processed to extract features, then a pass/reject decision is taken as the last step of the HLT workflow. Information from the full detector, such as the missing transverse momentum, may be required for the reconstruction of physics objects [211].

Data accepted by the HLT are sent to permanent storage by the Sub-Farm-Output (SFO). These events, persisted at CERN's Tier-0 data centre, are used for the full event reconstruction.

The aforementioned algorithms for feature extraction implemented in the HLT, together with the outcome of the L1 trigger, form the so-called *trigger chains*. Each chain is specialized for the detection of distinct physics signatures (e.g. leptons, photons, jets, missing transverse and total energy). A set of trigger chains characterizing a certain data-taking is referred to as *trigger menu*. Based on each trigger menu, events are then written to *streams*. The following streams exist:

- **Physics stream.** Events with complete detector information and interesting for all physics analyses. This is the most resource-intensive stream for computing and storage requirements;

- **Express stream.** This is a reduced subset of the previous stream. It contains the events reconstructed offline in real time and used for monitoring and data quality checks;
- **Debug stream.** This stream contains the events for which no trigger decision could be taken and thus, they need to be analysed and recovered for fixing potential issues in the TDAQ;
- **Calibration stream.** This set comprises events triggered by algorithms focussing on specific sub-detectors. Only partial detector information is stored using the Partial Event Building (PEB);
- **Monitoring stream.** The events of this stream are not recorded but they are sent to dedicated nodes to be analysed online for monitoring purposes;
- **Trigger Level Analysis stream.** This stream includes events with only partial detector information and specific physics objects reconstructed by the HLT that are directly used by the Trigger Level Analysis.

As discussed earlier, the only way to overcome the large SM background at low masses is to collect data with high statistics. The reduced event size of the TLA stream allows to significantly increase the event rate while remaining within the limits of the available resources. In fact, the *bandwidth*, i.e. the amount of data that can be transferred per unit time, is the limiting factor for the TLA data taking.

Similarly to what has been presented in Chapter 4, it is usually difficult, if possible at all, to handle large amounts of data at high speed over large distances. Therefore, the first trigger level (L1), which must be able to handle an event rate of up to 40 MHz, is built on dedicated hardware and located in the closest vicinity of the actual detector's readout. Instead, the High Level Trigger (HLT) can only accept event rates of up to 100 kHz and this represents a limitation to the output event rate of level L1. Lastly, the long-term storage can only cope with event rates of up to ~ 1 kHz, which, in turn, represents a limitation on the output of the HLT.

As a general indication, inclusive triggers useful to many analyses are recorded in the main stream characterized by 1 MB events with an average rate of 1 kHz. Dedicated triggers for flavour physics streams are usually processed by custom reconstruction algorithms and, despite the same event size of 1 MB, can only achieve event rates of the order of 200 Hz.

The solution adopted by the Trigger Level Analysis is to reduce the event size down to ~ 5 kB in order to acquire data with better statistics thanks to the increased event rate, which can be as high as 26 kHz. Smaller-sized events are retrieved at the HLT level by

omitting the raw detector output data and the fully calibrated and reconstructed *offline* objects.

As shown in Fig. 8.2, the increased collected statistics results in a better sensitivity in the low-mass regions, where the event rate allowed by the traditional trigger algorithms would be a strong limitation. In particular, the black dots show how, without any pre-scaling, TLA can benefit from the same statistics as the pre-scaled data (after being corrected by the pre-scale factor). Or, equivalently, the smaller-sized events used by TLA allow *online* data collection with comparable statistical power as the reconstructed *offline* events.

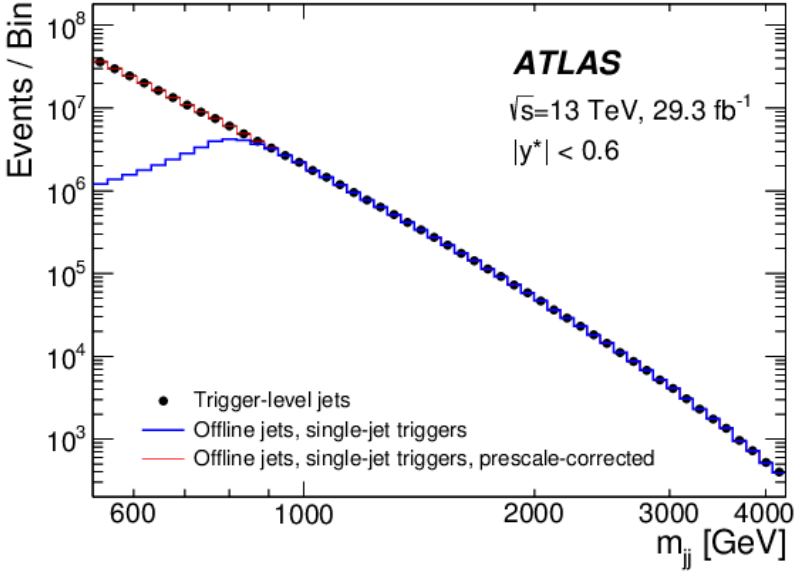


Figure 8.2: Comparison between the number of di-jet events in the data (black dots), the number of events selected by a single-jet trigger (thicker, blue line), and the number of events selected by single-jet triggers once corrected for the trigger prescale factors (thinner, red line) as a function of the di-jet invariant mass (m_{jj}) [205].

8.0.2 Data samples

The TLA uses topo-cluster objects reconstructed at the HLT level using the anti- k_t algorithm (Sec. 7.2.1) with a radius parameter $R = 0.4$. In particular, the crucial informations available for the jet reconstruction process is:

- the four-momentum of each jet;

- a set of calorimeter variables which provide information about the structure of the jet.

Due to the limitations on the event size discussed earlier, tracking information cannot be included.

In this study, TLA selects and records only Run2 events (only those collected in 2016) containing at least one jet with $E_T > 100$ GeV with a total luminosity of 29.3 fb^{-1} .

An additional subset of data is obtained requiring at least one jet with $E_T > 75$ GeV yielding a luminosity of 3.6 fb^{-1} .

8.0.3 Calibration Procedure

The main steps of the offline jet calibration process are summarized as follows:

1. Pile-up is subtracted based on the area of the jet;
2. Pile-up offset correction as a function of the Number of Primary Vertices (NPV) and of the average interactions per bunch crossing (μ) is derived from Monte Carlo simulations;
3. The correction of the jet origin, which is supposed to start at the detector center is applied. This only affects jet position and mass;
4. Jet Energy Scale and position calibrations are derived starting from the electromagnetic scale. Correction factors are derived from Monte Carlo simulations;
5. Corrections based on tracking, calorimeter and muon chamber information are applied. This allows to improve the Jet Energy Resolution (JER) and this process is known as Global Sequential Calibration (GSC);
6. η -intercalibration of data is calculated from the ratio between data and MC considering a central and a more forward jet;
7. Similarly, data/MC ratios (based on in-situ techniques) are also calculated to balance jets with well-calibrated objects.

The same general approach used for offline jets is also applied for trigger jet calibration. However, since HLT jets are built exclusively from calorimeter data and lack tracking information, it is not possible to perform each single step of the offline calibration as shown in Fig. 8.3. Data-derived corrections in bins of η and p_T are applied to calibrate the trigger

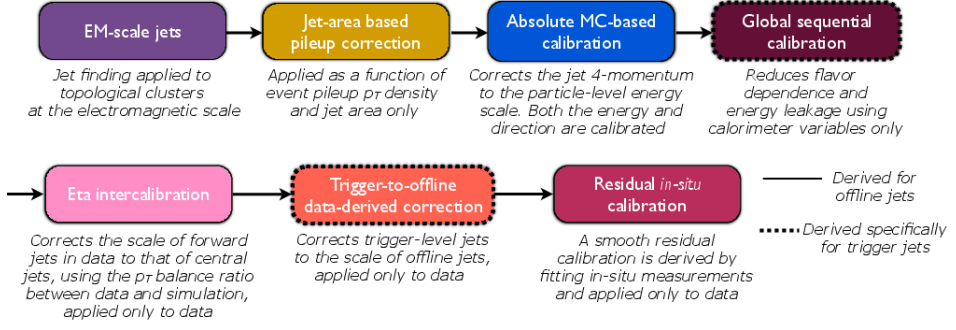


Figure 8.3: Calibration stages for EM-scale trigger-level jets. The boxes marked with a solid outline represent steps applied as derived for offline jets, the boxes with dashed outlines represent additional or modified steps specific to online jets [205].

jets to the same energy scale in order to match as closely as possible the properties of offline jets. The same in-situ calibration step and the same systematic uncertainties are considered for both types of jets.

8.o.4 Event selection

The TLA di-jet event selection requires that events must contain at least two trigger jets, each with a transverse momentum greater than 85 GeV and $|\eta| < 2.8$. The selection imposes that the hardest jet (i.e. the *leading* jet) in each event must have $p_T > 185$ GeV for the dataset with E_T higher than 75 GeV and $p_T > 220$ GeV for the dataset with E_T higher than 100 GeV.

The events containing jets seeded by non-collisional sources, such as calorimeter noise bursts, beam-induced background or cosmic rays, need to be rejected. In order to perform this rejection, TLA applies the criteria described in [212] or [213]. Since tracking information is not available, the selections based on tracks have been omitted. The resulting impact has been judged negligible.

Globally, the mass range that the TLA considers for the di-jet resonance searches extends from 450 GeV to 1800 GeV. Two distinct selection criteria are applied for two different and partially overlapping di-jets invariant mass regions:

- for $700 \text{ GeV} < m_{jj} < 1800 \text{ GeV}$ mass range, events need to satisfy $|y^*| < 0.6$;
- for $450 \text{ GeV} < m_{jj} < 1800 \text{ GeV}$ mass range, events need to satisfy $|y^*| < 0.3$; the latter criterion is applied to the smaller dataset characterized by at least one L1 jet with $E_T > 75 \text{ GeV}$.

8.0.5 Background estimation

The Standard Model predicts the QCD background for TLA searches in the di-jet mass distribution to be smooth and falling, whereas a new BSM state decaying into two partons would emerge as a localized excess in such distribution. In order to perform a background estimation, TLA adopts a data-driven fit to the di-jet mass distribution: smooth functional forms that do not adapt to localized excesses are used for the fit. With this method, which relies on the absence of unexpected features that may arise from detector malfunctioning or calibration, background systematic uncertainties are minimized.

To obtain the results presented in this section, the SWiFt approach is used to determine the SM background distribution (see Sec. 7.3.3). As discussed in [205], two windows of 19 and 27 bins have been used to evaluate the background for m_{jj} spectra starting from 531 GeV and 400 GeV, respectively. The width of the windows for the SWiFt analysis is optimized to produce the best χ^2 and, at the same time, to ensure that the fit remains insensitive to potential signal contributions.

Three functional forms have been used for each sliding window: a five-parameter function (Eq. 8.1), a four-parameter function obtained from Eq. 8.1 by setting $p_5 = 0$ and a four-parameter function (Eq. 8.2).

$$f(x) = p_1(1 - x)^{p_2} x^{p_3 + p_4 \ln x + p_5 \ln x^2} \quad (8.1)$$

$$f(x) = \frac{p_1}{x^{p_2}} e^{-p_3 x - p_4 x^2} \quad (8.2)$$

where p_i are the free parameters and $x = m_{jj}/\sqrt{s}$.

8.0.6 Search phase results

The TLA's objective is to find evidences of new physics in the form of a localized excess of events.

Fig. 8.4, represents the experimental invariant mass distribution that has been fitted with the background-only-hypothesis; the functions presented in the previous section have been considered and, in this case, the sliding window technique has been applied. In order to identify the most discrepant interval, the BumpHunter algorithm [203] has been run and the most significant discrepancy has been found in the region between 889 and 1007 GeV (blue dashed vertical lines).

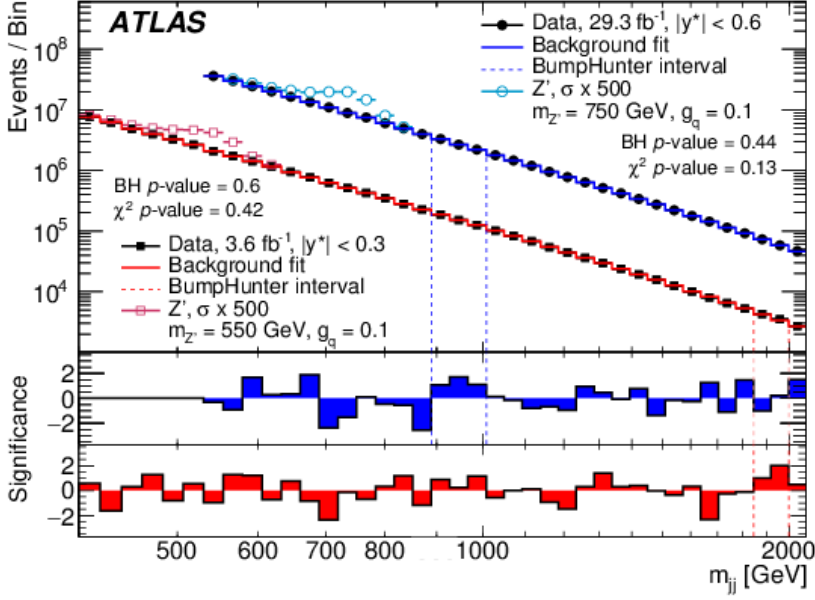


Figure 8.4: Reconstructed di-jet mass distribution (filled points) for events in the $|y^*| < 0.3$ and $|y^*| < 0.6$ signal regions. The solid lines represent the background estimate from a sliding-window fit. The χ^2 p -value quantifies the agreement between the background estimate and the data. Vertical lines identify the largest excesses found by the BumpHunter. The circles show two possible signal models. The lower panels show the bin-by-bin significances, considering only statistical uncertainties [205].

The probability to observe a deviation, considering the entire range of the mass distribution at least as significant as that observed in data is 0.44; this value leads to the conclusion that there is no evidence of any local excess¹.

Following the procedure reported in Sec. 7.3, if statistically significant evidence for new physics in the di-jet spectrum does not emerge during the search phase, it is possible to proceed with the limit setting.

8.o.7 Limit setting

In general, the limit setting phase aims to evaluate the probability that a particle with a given mass and behaviour exists, given the observed data.

¹A p -value equivalent to 0.44 corresponds to a significance of 0.16σ [200]. Often, in HEP, a discovery is claimed when the p -value of the background only hypothesis is found below $2.9 \cdot 10^{-7}$, i.e. 5σ effect [200].

In the limit setting step, the attention has been focused both on a model-dependent case, the leptophobic Z' simplified model, and on a generic template shape, a Gaussian distribution able to emulate the behaviour of abstract resonances.

Leptophobic Z' simplified model In this case, since the model assumes that Z' is leptophobic, only an axial-vector interaction with quarks is considered and no interactions with leptons ($g_l = 0$).

The signal samples have been generated assuming that the decay rate of the Z' mediator into DM final states is negligible; this makes the di-jet production rate and the resonance width only dependent on two free parameters: the flavour-universal coupling g_q of Z' to quarks and the mass of the resonance $m_{Z'}$.

To produce signal samples, the matrix elements were calculated using MADGRAPH 5 [214] while the parton showering processes were simulated using PYTHIA8 [215].

As shown in Fig. 8.5, limits on the product of the cross-section (σ), the acceptance² (A) and the model branching ratio (B) have been set. The results are presented in the $g_q, m_{Z'}$ plane, thus transforming the above-mentioned limits into limits on g_q . This procedure was done considering the scaling of the signal cross section with g_q^2 .

Generic Gaussian models Generic Gaussian models represent a class of paradigms where the DM signal is modelled as a Gaussian contribution to the observed m_{jj} spectrum. These models, although more generic than the previous one, can be effectively adopted for a broad set of models characterized by Gaussian-like signals.

This approach can be applied if:

- the resonances are sufficiently narrow and exhibit a Gaussian shape near the peak with tails that are much smaller than background;
- the impact of the parton distribution functions and non-perturbative effects can be overlooked.

In this study, four different Gaussian signal widths are considered for each mass value taken into account: 5%, 7% and 10% and a width value equal to the simulated mass resolution

²The value of the acceptance can vary depending on the model adopted, the mass value and $|y^*|$; as an example, some typical acceptance values adopted in this search are listed below:

- for a signal mass of 550 GeV, $g_q = 0.10$, $|y^*| < 0.3$, the acceptance considered is 20%;
- for a signal mass of 750 GeV, $g_q = 0.10$, $|y^*| < 0.6$, the acceptance considered is 41%.

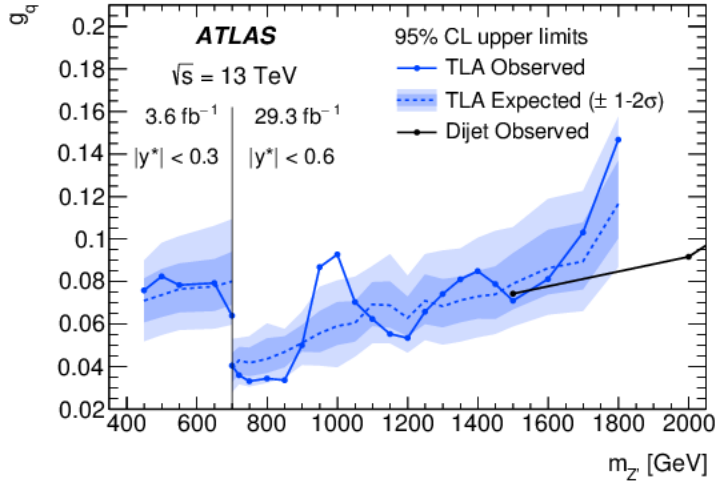


Figure 8.5: The 95% confidence-level observed and expected upper limits on g_q as a function of m_Z . Couplings above the solid lines are excluded. The solid and dashed lines represent the observed and expected limits, respectively [205].

which can vary between 4% and 6%. The increasing widths are tested to distribute the expected signal contribution across different bins and thus minimize the effect that statistical fluctuations from the data can have on the signal. In Fig. 8.6, the limits on $\sigma \times A \times B$ for a hypothetical signal producing a Gaussian contribution to the observed m_{jj} are shown.

In the process of setting limits on new phenomena, the effects of several systematic uncertainties must be considered. As explained in Sec. 7.3, it is possible to identify uncertainties related to the background estimation and uncertainties relative to the signal. The Trigger Level Analysis adopts a data-driven technique to perform the background estimation considering two independent background uncertainties:

- uncertainty due to the fit function choice;
- uncertainty due to the fit function parameters.

To assess the effect of the functional form choice, the fit results from a *nominal* and an *alternative* function are compared. In particular:

- for each signal selection, the chosen function is the one yielding the best χ^2 over the full fitted m_{jj} range;

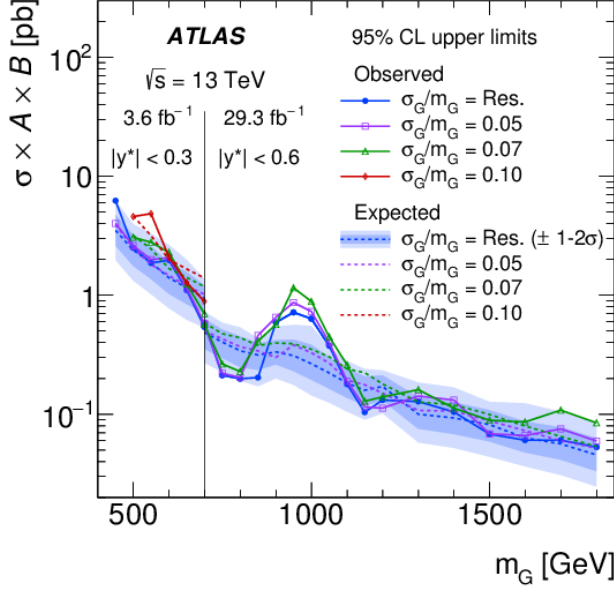


Figure 8.6: The 95% confidence-level observed upper limits on $\sigma \times A \times B$ for two jets for a hypothetical Gaussian signal contribution to the observed m_{jj} distribution. The limits are shown for a relative width σ_G/m_G corresponding to a width equal to the detector mass resolution [205].

- for the signal selection with $|y^*| < 0.6$, Eq. 8.1 is used as the nominal function and its four-parameter version (i.e. with $p_5 = 0$) represents the alternative choice;
- for the signal selection with $|y^*| < 0.3$, the four-parameter version of Eq. 8.1 ($p_5 = 0$) is used as the nominal function, whereas Eq. 8.2 is used as the alternative choice.

The uncertainties on the fit parameter values are obtained from pseudo-experiments by drawing pseudo-data from Poisson-fluctuated values of the nominal background model. These two systematics, related to the background estimation, have been included for both the studied signals.

The uncertainties on the Z' signal model include the Jet Energy Scale and the luminosity, while the impact of the Jet Energy Resolution is negligible. In accordance with the measured impact on the Z' samples, a constant JES uncertainty of 3% is applied for the Gaussian model. The integrated luminosity uncertainty is $\pm 2.2\%$.

For the observed limits shown in Fig. 8.5 and Fig. 8.6, the Bayesian approach to limit calculation has been adopted [216].

The observed number of events N in a given m_{jj} spectrum is represented as the sum of a signal S and a background B contributions (Eq. 8.3):

$$N = \mu S + B \quad (8.3)$$

where:

- S is the nominal number of signal events predicted by the signal model under investigation;
- B is the number of background events according to the SM predictions;
- μ is a coefficient representing the *signal strength*.

The signal+background model being tested against measured data does not depend uniquely on the parameter of interest μ . Instead, it must take into account all the possible uncertainties and sources of error. These are represented by the so-called *nuisance parameters* $\theta = \theta_1, \theta_2, \dots$. However, to obtain a mono-dimensional probability density function depending only on the signal strength, it is necessary to integrate the multidimensional Probability Density Function (PDF) over all the nuisance parameters, as shown in Eq. 8.4. This step is referred to as *marginalization* and is the core of the Bayesian approach to limit calculation.

$$p(\mu|x) = \int p(\mu, \theta|x) d\theta \quad (8.4)$$

According to Bayes' theorem and considering the independence of the signal strength from all the nuisance parameters considered by this analysis, Eq. 8.4 can be rewritten as Eq. 8.5.

$$p(\mu|x) \propto \int \mathcal{L}(x|\mu, \theta) \pi(\mu) \prod_i \pi(\theta_i) d\theta \quad (8.5)$$

The previously defined uncertainties are included into the Bayesian marginalization with Gaussian priors ($\pi(\theta_i)$). A single nuisance parameter is defined for each uncertainty.

Expected limits are calculated from pseudo-experiments by taking into account all the possible uncertainties. When provided together with observed upper limits, they allow to quantify the number of standard deviations between the observed limit and what would be expected at most by the sole null hypothesis. In ATLAS, expected limits are calculated with

a frequentist approach based on pseudo-datasets generated from a chosen background-only hypothesis.

For masses of approximately 1 TeV in the $|y^*| < 0.6$ signal region, a previously unobserved (Fig. 8.4) upward mass fluctuation is present. A background-only hypothesis is used for the sliding window fit during the excess search phase. Instead, for the expected and observed limits, in addition to the background parameterization, a signal shape is also included to adapt to local data fluctuations. The difference in these two approaches is particularly pronounced in the $|y^*| < 0.6$ region. Signals with 10% widths in the $|y^*| < 0.6$ region are too wide for the sliding window size and have been excluded from the limit setting phase.

8.0.8 Current status

The search for di-jet resonances in the mass range between 450 GeV and 1800 GeV using trigger-level jets in 29.3 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ proton-proton collision data has not resulted in significant localized excesses over the estimated SM background. 95% confidence level limits on Z signals and cross-sections producing Gaussian contributions to the di-jet mass spectra are provided. For a large range of masses, the sensitivity to the quark coupling g_q is doubled with respect to pre-LHC and to $\sqrt{s} = 8$ and 13 TeV ATLAS results. Gaussian contributions with effective cross-sections times acceptance in the range between 6.5 pb at 450 GeV to 0.4 pb at 700 GeV, to 0.05 pb at 1800 GeV are excluded.

8.1 Improvements to the fitting procedure using Frequentist tools

The limits on the cross-section discussed in the previous section have been calculated with a Bayesian approach.

However, several analysis groups in ATLAS adopt a frequentist approach and, in addition, the comparison between TLA (ATLAS) limits and the results obtained by the corresponding analyses in CMS, has highlighted non-negligible discrepancies deserving further investigation [217]:

- Low mass (800 GeV) - $3\times$ difference between ATLAS and CMS for expected limits;
- High mass (1.4 TeV) - $2\times$ difference between ATLAS and CMS for expected limits.

Since CMS also adopts a frequentist approach, the observed differences have been ascribed to the inherent differences in the two statistical approaches and to the different treatment of uncertainties on the fit parameters in each.

These considerations led the TLA group to start a preliminary study to assess the possibility of replacing the Bayesian approach with a frequentist procedure during the limit setting phase.

In order to identify the most suitable tool for the TLA requirements, several frequentist tools have been included in the preliminary investigation presented in this section:

- HistFitter [218];
- xmlAnaWSBuilder [219];
- quickFit [220].

Although a conclusive statement would require an extensive analysis of all the aspects of these tools, the most relevant observations will be discussed. 95% confidence-level upper limits have been set using the HistFitter code (Sec. 8.1.1) and the result has been compared to the values obtained from the BayesianFramework tool; further investigations on the fit parameter errors have been carried out based on the xmlAnaWSBuilder and the quickFit tools (Sec. 8.1.2).

8.1.1 Setting limits with HistFitter

As discussed earlier, in absence of any excess, the observed data and the predicted background are used to set limits on the signal strength of new phenomena.

For the tool validation, the results obtained from the HistFitter have been compared to those obtained from the usual BayesianFramework tool. The limits found are summarized in Tab. 8.1³.

A generic Gaussian signal contribution to the observed m_{jj} distribution was considered. The values presented in Tab. 8.1 are relative to a signal centered at $m_{jj} = 800$ GeV with a width of 7% (i.e. $\pm 1\sigma$ -width at 744 and 856 GeV, respectively).

The dataset used for the limit calculation was collected at an integrated luminosity of up to 29.3 fb^{-1} and a centre of mass energy of 13 TeV. The figures in Tab. 8.1 were calculated from a fit performed on the mass range $531 < m_{jj} < 2081$ GeV.

As represented, the limits are calculated including the systematic uncertainties on the background that the TLA group is currently taking into account:

- uncertainty relative to the choice of function for the fit;

³The values are not multiplied by cross-section, acceptance and Branching Ratio (BR).

Table 8.1: Comparison between TLA limits obtained with the reference BayesianFramework tool and with the HistFitter.

	BayesianFramework	HistFitter
No systematics	6118.52	6117.71
Fit function choice	6118.54	6117.51
Fit parameter values	6257.82	6257.71

- uncertainty associated to the errors on the fit parameter values.

These preliminary studies show that the two statistical methods set limits which are compatible within 1%.

Despite the overall agreement, the HistFitter has not been confirmed by the TLA group as the frequentist tool of choice. Indeed, its implementation does not allow to explicitly provide analytical representations of the PDFs to be used during the fitting phase. Instead, an automatic and integrated procedure is applied [221]. However, the quality of the fit is not consistent throughout the mass range and, for some mass values, the convergence is not reached.

8.1.2 Review and improvements of the fitting procedure

The purpose of this study is to improve the stability and quality of a background-only fit. The investigations were carried out using RooFit-based frequentist tools, namely the `xmlAnaWSBuilder` and the `quickFit` codes, as compared to the reference Bayesian framework. Two aspects have been considered:

- the role of the p_1 parameter (Eq. 8.1 or 8.2);
- the fit quality and the behaviour of the statistical tools.

The role of the normalization factor p_1 The workflow under investigation entails a preliminary step carried out with the `xmlAnaWSBuilder`, then followed by the actual fit performed by the `quickFit` code.

The first stage is responsible of creating the RooFit workspace starting from the ROOT datasets and from the analytical definitions of the functions.

However, the `xmlAnaWSBuilder` is also capable of performing simple fits. When this option is enabled, it is observed that the value of p_1 doubles at each fit pass. It is important

to stress that all the other parameters remain unaffected. It is therefore necessary to ensure that no successive fits are performed on the same datasets.

As explained in [221], in frequentist tools, the handling of the PDF normalization is independent from the definition of a normalization parameter, which would be the role of p_1 in Eq. 8.1 or 8.2. Therefore, to avoid the wrong treatment of the function normalization, it is necessary to remove p_1 (or, alternatively, to fix its value to 1) from the analytical definition of the PDF.

With the aforementioned precautions, the fit stability and reproducibility is improved.

Fit parameters uncertainties Unlike the Bayesian framework, in RooFit-based codes, the estimate of the fit parameter errors is based on the calculation of the correlation matrix (Sec. 7.3.5), which in turn depends on the covariance matrix. The ij -th element of the correlation matrix is calculated according to Eq. 8.6 [222].

$$C_{ij} = \frac{\text{Cov}(i, j)}{\sqrt{\text{Cov}(i, i)\text{Cov}(j, j)}} \quad (8.6)$$

where i and j represent the PDF parameters.

Once the correlation matrix C is available, it can be combined with the *vector of variations* $F(x)$ to obtain the error $E(x)$ according to Eq. 8.7.

$$E(x) = F(x)CF(x) \quad (8.7)$$

The vector of variations is defined according to Eq. 8.8

$$F_i(x) = \frac{1}{2} \left(\text{PDF}(x, \vec{p}) \Big|_{p_{i,\text{fit}} + Z \cdot e_{i,\text{fit}}; p_{j \neq i}} - \text{PDF}(x, \vec{p}) \Big|_{p_{i,\text{fit}} - Z \cdot e_{i,\text{fit}}; p_{j \neq i}} \right) \quad (8.8)$$

A simplified visualization of this calculation is shown in Fig. 8.7.

The upper (lower) curve, represented in Fig. 8.7 with the label 2 (3), is created for each parameter i by evaluating the nominal PDF (label 1) with the i -th parameter set to the value obtained from the fit plus (minus) Z times the error on the parameter itself, as calculated by the fit. This is done for each individual parameter i , while keeping all the other parameters j fixed to the nominal values (Z is the width in number of sigmas at which the error is evaluated). The components $F_i(x)$ of the vector of variations correspond to the error band indicated with label 5 in Fig. 8.7.

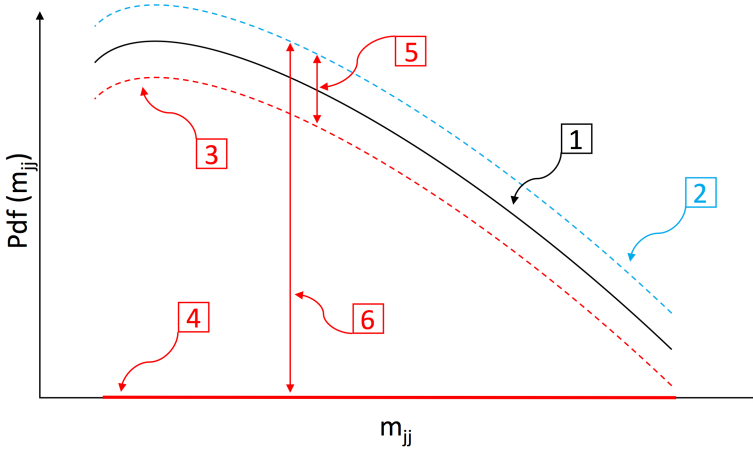


Figure 8.7: Simplified representation of the error-band calculated from the correlation matrix.

RooFit, the ROOT toolkit for statistical analysis, has the built-in functionality to calculate the correlation and covariance matrices in order to evaluate the fit parameter errors as described above. This capability is provided by the `PlotOn` visualization routine and, in particular, by its `VisualizeErrors` option. The user can retrieve the values of the calculated errors used to produce the error bands.

The following workflow has been adopted for the study:

1. First, the analytical definition of the background PDF is provided to the RooFit-based `xmlAnaWSBuilder` tool. At this stage, nominal values and ranges for the parameters are defined. The Parameter of Interest (POI) and the nuisance parameters are also identified. This allows the `xmlAnaWSBuilder` to create a RooFit Workspace;
2. The RooFit workspace is then passed to the `quickFit` tool, which performs the fit to data using the PDF previously defined. The result of this second step is a RooFit object, called `RooFitResult`, containing the parameter values and all the fit information to build the covariance matrix;
3. Lastly, the initial workspace, providing the parameters' nominal values, and the post-fit result file are passed to the `PlotOn` method for visualization. As anticipated, the estimate of the error on the fit parameters is performed at this stage when the `VisualizeErrors` option is called.

It is worth focusing the attention on the behaviour of the RooFit algorithm. When the fit function is defined, each parameter is initialized with a nominal value and a range within which it is allowed to float. After the fit is completed, each parameter is associated to an error and, from this, the correlation matrix is obtained. The vector of variations is calculated by adding (subtracting) the error to the parameter. If the resulting parameter value is larger (smaller) than the maximum (minimum) value allowed by the parameter's range, then the parameter value is set to the corresponding range boundary. In case the lower curve results in a negative-valued function, hence not representing an actual PDF, then the whole curve is set to 0 (label 4 in Fig. 8.7). The resulting error bar is therefore largely overestimated (label 6 in Fig. 8.7).

The major implication of this behaviour on the estimation of the fit parameter error is the dependency of the error itself on the prior choice of the parameter range. A large error combined with a small range results in an underestimation of the actual error, because the PDF values are set to the range boundaries. Rather than being dominated by the parameter error itself, the overall uncertainty on the fit is driven by the initial choice of parameter intervals. Additionally, unless properly initialized, a strong dependency of the fit parameter errors on the choice of parameter range has also been observed for the quickFit tool. This tool is therefore not suitable for an explorative fit phase. To overcome this limitation, a ROOT-based pre-fit script has been developed [223].

As shown in Fig. 8.8, the quality of the background fit performed with quickFit is dramatically improved when the pre-fit stage is applied.

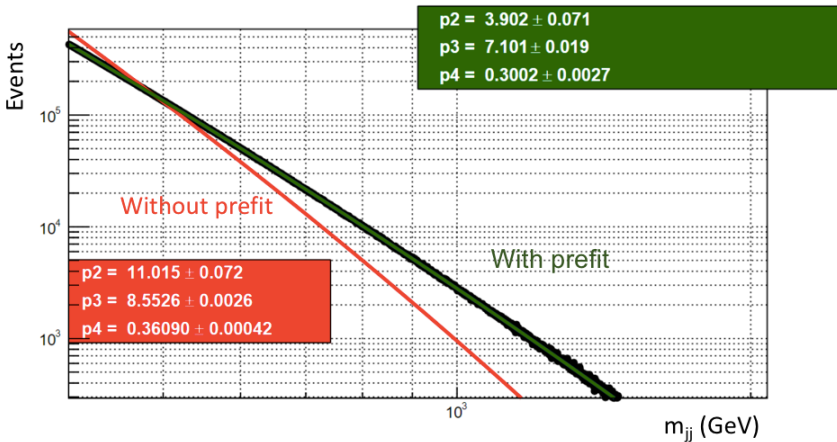


Figure 8.8: Background fit obtained from quickFit with (green curve) and without (red curve) the ROOT pre-fit.

8.2 Conclusions and outlook

TLA searches are characterized by many challenging aspects: the high QCD background, the very small signature signals of DM particles, the interest in the lower mass part of the spectrum, where the event rates are limited by the trigger thresholds, just to name a few. In addition, the use of events with reduced information affects the jet energy scale, the resolution and the calibration: these should be derived from offline jets data, but the pre-scaling significantly reduces the statistical power of this information.

The replacement of Bayesian tools with their frequentist counterparts is not straightforward. Some tools, such as the HistFitter, are characterized by an automated process for building PDFs, thereby restricting the user's ability to input specific functional forms. The other tools tested have shown a strong sensitivity to the choice of nominal values and ranges for the fit parameters; in fact, it is recommended to perform a simple explorative pre-fit before running the actual fit, in order to restrict the ranges within which parameters can float.

However, this preliminary investigation is not conclusive and it is strongly recommended to assess the robustness of the FrequentistFramework, and to take into account a more comprehensive set of frequentist codes, which could provide specific tools to handle the initial explorative fit phase. In addition, more tests can be carried out on the tool presented in this thesis, such as the possibility of performing a SWiFt analysis to further improve the quality of the background-only fit.

Concluding remarks

As thoroughly reviewed in the opening chapters of this work, extending the Standard Model with the discovery of Dark Matter entails a huge effort from the entire scientific community. The elusive nature of DM introduces an unprecedented level of complexity in the design of suitable experiments. Higher energies, higher luminosities, higher computational power are only a few of the challenges currently being addressed.

This thesis deals with two of the main aspects of the planned upgrades of the LHC, intended to boost DM searches at colliders: the optimization of the existing analysis codes and the identification of statistical tools sensitive to weak signals and capable of increasingly higher statistical accuracy.

New opportunities have been found for the optimization of the Geant4 workflow and for improving the efficiency of the geometry description of the ATLAS detector. *In-vitro* studies carried out on standalone benchmark tools have shown promising results and it is highly recommended to continue and expand the optimization effort in order to eventually reach the Grid scale.

In addition, the preliminary studies on frequentist tools applied to the ATLAS TLA analysis identified the possibility of an improvement in the fit quality and reproducibility. This will eventually result in a more accurate procedure of limit setting and, possibly, even to the discovery of the elusive Dark Matter.

It is left to future developments of this work to upscale the research to the extent of the HL-LHC upgrades.

Bibliography

- [1] W. Heisenberg and I. Paggi. *Fisica e oltre. Incontri con i protagonisti (1920-1965)*. Universale Bollati Boringhieri-S. scient. Bollati Boringhieri, 2008. ISBN: 9788833918655.
- [2] G. Reale and V. Cicero. *Storia della filosofia greca e romana*. Bompiani, 2018. ISBN: 9788858780954.
- [3] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. DOI: 10.1017/Cbo9781139525367.
- [4] B. Abbott et al. “First upper limits from LIGO on gravitational wave bursts”. In: *Phys. Rev. D* 69.10 (2004), page 102001.
- [5] B. Abbott et al. “Analysis of LIGO data for gravitational waves from binary neutron stars”. In: *Phys. Rev. D* 69.12 (2004), page 122001.
- [6] G. Kane. *Modern Elementary Particle Physics: Explaining and Extending the Standard Model*. 2nd edition. Cambridge University Press, 2017. ISBN: 9781107165083. DOI: 10.1017/9781316691434.
- [7] D. Griffiths. *Introduction to Elementary Particles*. Wiley-VCH Verlag GmbH, 2008. ISBN: 9780471603863. DOI: 10.1002/9783527618460.
- [8] W. N. Cottingham and D. A. Greenwood. *An Introduction to the Standard Model of Particle Physics*. 2nd edition. Cambridge University Press, 2007. ISBN: 9780521852494. DOI: 10.1017/Cbo9780511791406.
- [9] J. Woithe, G. J. Wiener, and F. F. Van der Veken. “Let’s have a coffee with the Standard Model of particle physics”. In: 52.3 (2017), page 034001. ISSN: 1361-6552. DOI: 10.1088/1361-6552/aa5b25.
- [10] F. Strocchi. “Simmetrie e rottura di simmetrie in Fisica”. In: ().
- [11] V. V. Belokurov, D. V. Shirov, and P. Millard. *The Theory of Particle Interactions*. American Institute of Physics translation series. American Inst. of Physics, 1991. ISBN: 9780883187159.
- [12] F. Halzen and A. D. Martin. *Quarks And Leptons: An Introductory Course In Modern Particle Physics*. 1984. ISBN: 978-0-471-88741-6.

- [13] D. H. Perkins. *Introduction to High Energy Physics*. 4th edition. Cambridge: Cambridge University Press, 2000. ISBN: 9780521621960. DOI: 10.1017/Cbo9780511809040.
- [14] B. Kayser. “Neutrino Mass, Mixing, And Oscillation”. In: *Flavor Physics for the Millennium*. World Scientific, 2001.
- [15] Zyla et al. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2020.8 (2020).
- [16] H. Fritzsch. “The history of QCD”. In: *CERN Courier* 52.8 (2012), pages 21–25.
- [17] F. Abe et al. “Observation of Top Quark Production in $verline\backslash mathitp\backslash mathitp$ Collisions with the Collider Detector at Fermilab”. In: *Phys. Rev. Lett.* 74.14 (1995), pages 2626–2631.
- [18] S. Abachi et al. “Observation of the Top Quark”. In: *Physical Review Letters* 74.14 (1995), pages 2632–2637.
- [19] S. K. Choi et al. “Observation of a Narrow Charmoniumlike State in Exclusive $B^\pm \rightarrow K^\pm \pi^+ \pi^- J/\psi$ Decays”. In: *Phys. Rev. Lett.* 91.26 (2003), page 262001.
- [20] R. Aaij et al. “Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_- b^0 \rightarrow J/\psi K^- p$ Decays”. In: *Phys. Rev. Lett.* 115.7 (2015), page 072001.
- [21] P. A. M. Dirac and N. H. D. Bohr. “The quantum theory of the emission and absorption of radiation”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 114.767 (1927), pages 243–265. DOI: 10.1098/rspa.1927.0039.
- [22] E. Fermi. “Versuch einer Theorie der β -Strahlen. I”. In: *Zeitschrift für Physik* 88.3 (1934), pages 161–177. ISSN: 0044-3328. DOI: 10.1007/Bf01351864.
- [23] S. L. Glashow. “The renormalizability of vector meson interactions”. In: *Nuclear Physics* 10 (1959), pages 107–117. ISSN: 0029-5582. DOI: 10.1016/0029-5582(59)90196-8.
- [24] A. Salam and J. C. Ward. “Weak and electromagnetic interactions”. In: *Il Nuovo Cimento (1955-1965)* 11.4 (1959), pages 568–577. ISSN: 1827-6121. DOI: 10.1007/Bf02726525.
- [25] S. Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19.21 (1967), pages 1264–1266.
- [26] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13.16 (1964), pages 508–509.
- [27] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13.9 (1964), pages 321–323.
- [28] P. Skands. “Introduction to QCD”. In: *Searching for New Physics at Small and Large Scales* (2013). DOI: 10.1142/9789814525220_0008.

- [29] S. Höche. *Introduction to parton-shower event generators*. 2015.
- [30] B. R. Webber. “A QCD model for jet fragmentation including soft gluon interference”. In: *Nuclear Physics B* 238.3 (1984), pages 492–528. ISSN: 0550-3213. DOI: 10.1016/0550-3213(84)90333-X.
- [31] B. R. Webber. “Fragmentation and Hadronization”. In: (1999).
- [32] B. Andersson et al. “Parton fragmentation and string dynamics”. In: *Physics Reports* 97.2 (1983), pages 31–145. ISSN: 0370-1573. DOI: 10.1016/0370-1573(83)90080-7.
- [33] G. Sterman and S. Weinberg. “Jets from Quantum Chromodynamics”. In: *Phys. Rev. Lett.* 39.23 (1977), pages 1436–1439.
- [34] G. Sterman. *QCD and Jets*. 2004.
- [35] J. D. Bjorken. “Asymptotic Sum Rules at Infinite Momentum”. In: *Phys. Rev.* 179.5 (1969), pages 1547–1553.
- [36] J. Campbell, J. Huston, and F. Krauss. *The Black Book of Quantum Chromodynamics*. Oxford University Press, 2018.
- [37] *The Durham HepData Project*. URL: <http://hepdata.cedar.ac.uk/pdf/pdf3.html>.
- [38] G. Altarelli and G. Parisi. “Asymptotic freedom in parton language”. In: *Nuclear Physics B* 126.2 (1977), pages 298–318. ISSN: 0550-3213. DOI: 10.1016/0550-3213(77)90384-4.
- [39] J. Winter et al. *Systematic improvement of QCD parton showers*. 2012.
- [40] C. Burgess and G. Moore. *The Standard Model: A Primer*. Cambridge University Press, 2006. ISBN: 9780521860369. DOI: 10.1017/Cbo9780511819698.
- [41] A. Secchi. *L’Astronomia in Roma nel pontificato di Pio IX: memoria*. Tipografia della Pace, 1877.
- [42] G. Bertone and D. Hooper. “History of dark matter”. In: *Rev. Mod. Phys.* 90.4 (2018), page 045002.
- [43] F. Zwicky. “On the Masses of Nebulae and of Clusters of Nebulae”. In: *The Astrophysical Journal* 86 (1937), page 217. DOI: 10.1086/143864.
- [44] K. C. Freeman. “On the Disks of Spiral and So Galaxies”. In: *The Astrophysical Journal* 160 (1970), page 811. DOI: 10.1086/150474.
- [45] D. H. Rogstad and G. S. Shostak. “Gross Properties of Five Scd Galaxies as Determined from 21-CENTIMETER Observations”. In: *The Astrophysical Journal* 176 (1972), page 315.

- [46] Asher Yahalom. “The Effect of Retardation on Galactic Rotation Curves”. In: *Journal of Physics: Conference Series* 1239 (May 2019), page 012006. DOI: 10.1088/1742-6596/1239/1/012006. URL: <https://doi.org/10.1088/1742-6596/1239/1/012006>.
- [47] *What is Gravitational Lensing?* URL: <https://www.universetoday.com/118751/what-is-gravitational-lensing/>.
- [48] *Hubble sees dark matter ring in a galaxy cluster*. URL: https://www.esa.int/Science_Exploration/Space_Science/Hubble_sees_dark_matter_ring_in_a_galaxy_cluster.
- [49] M. J. Jee et al. “Discovery of a Ringlike Dark Matter Structure in the Core of the Galaxy Cluster Cl 0024+17”. In: *The Astrophysical Journal* 661.2 (2007), pages 728–749. DOI: 10.1086/517498.
- [50] *NASA Finds Direct Proof of Dark Matter*. URL: http://www.nasa.gov/home/hqnews/2006/aug/HQ_06297_CHANDRA_Dark_Matter.html.
- [51] *Searching for Dark Matter with the ATLAS detector*. URL: <https://atlas.cern/updates/feature/dark-matter>.
- [52] *Introduction: The Millennium Simulation*. URL: <https://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium/>.
- [53] A. Liddle. *An Introduction to Modern Cosmology*. Wiley, 2015. ISBN: 9781118690253.
- [54] M. Raggi, V. Kozhuharov, and P. Valente. “The PADME experiment at LNF”. In: (2015).
- [55] G. Gamow. “The Origin of Elements and the Separation of Galaxies”. In: *Phys. Rev.* 74.4 (1948), pages 505–506. DOI: 10.1103/PhysRev.74.505.2.
- [56] L. Brink. *Nobel Lectures in Physics (2006 – 2010)*. World Scientific, 2014.
- [57] M. Tanabashi et al. “Review of Particle Physics”. In: *Phys. Rev. D* 98.3 (2018).
- [58] Planck Collaboration et al. “Planck 2013 results. XV. CMB power spectra and likelihood”. In: *Astronomy & Astrophysics* 571 (2014). DOI: 10.1051/0004-6361/201321573.
- [59] *Planck’s power spectrum of temperature fluctuations in the Cosmic Microwave Background*. URL: <https://sci.esa.int/web/planck/-/51555-planck-power-spectrum-of-temperature-fluctuations-in-the-cosmic-microwave-background>.
- [60] N. Aghanim et al. “Planck 2018 results”. In: *Astronomy & Astrophysics* 641 (2020), page 6. DOI: 10.1051/0004-6361/201833910.
- [61] K. N. Abazajian et al. *Light Sterile Neutrinos: A White Paper*. 2012.

- [62] S. D. M. White, C. S. Frenk, and M. Davis. “Clustering in a neutrino-dominated universe”. In: *The Astrophysical Journal* 274 (1983), page 1. DOI: 10.1086/184139.
- [63] S. Dodelson and L. M. Widrow. “Sterile neutrinos as dark matter”. In: *Phys. Rev. Lett.* 72.1 (1994), pages 17–20.
- [64] T. M. Undagoitia and L. Rauch. “Dark matter direct-detection experiments”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.1 (2015), page 013001. DOI: 10.1088/0954-3899/43/1/013001.
- [65] K. Abazajian, G. M. Fuller, and M. Patel. “Sterile neutrino hot, warm, and cold dark matter”. In: *Phys. Rev. D* 64.2 (2001), page 023501. DOI: 10.1103/PhysRevD.64.023501.
- [66] D. V. Naumov. “The Sterile Neutrino: A short introduction”. In: *EPJ Web Conf.* 207 (2019). DOI: 10.1051/epjconf/201920704004.
- [67] R. D. Peccei and H. R. Quinn. “Constraints imposed by CP conservation in the presence of pseudoparticles”. In: *Phys. Rev. D* 16.6 (1977), pages 1791–1797. DOI: 10.1103/PhysRevD.16.1791.
- [68] R. D. Peccei and H. R. Quinn. “CP Conservation in the Presence of Pseudoparticles”. In: *Phys. Rev. Lett.* 38.25 (1977), pages 1440–1443.
- [69] P. J. E. Peebles. “Primordial Helium Abundance and the Primordial Fireball. II”. In: *The Astrophysical Journal* 146 (1966), page 542. DOI: 10.1086/148918.
- [70] P. J. E. Peebles. “Structure of the Coma Cluster of Galaxies”. In: *The Astronomical Journal* 75 (1970), page 13. DOI: 10.1086/110933.
- [71] J. E. Kim. “Weak-Interaction Singlet and Strong CP Invariance”. In: *Phys. Rev. Lett.* 43.2 (1979), pages 103–107.
- [72] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov. “Can confinement ensure natural CP invariance of strong interactions?”. In: *Nuclear Physics B* 166.3 (1980), pages 493–506. ISSN: 0550-3213. DOI: 10.1016/0550-3213(80)90209-6.
- [73] S. P. Martin. In: volume Volume 18. Advanced Series on Directions in High Energy Physics. World Scientific, 1998. Chapter A Supersymmetry Primer, pages 1–98. ISBN: 978-981-02-3553-6. DOI: 10.1142/9789812839657_0001.
- [74] M. Carena et al. “Invisible Z-boson decays at e^+e^- colliders”. In: *Phys. Rev. D* 68.11 (2003), page 113007. DOI: 10.1103/PhysRevD.68.113007.
- [75] G. Aad et al. “Constraints on new phenomena via Higgs boson couplings and invisible decays with the ATLAS detector”. In: *Journal of High Energy Physics* 2015.11 (2015), page 206. ISSN: 1029-8479. DOI: 10.1007/Jhep11(2015)206.
- [76] A. Boveia and C. Doglioni. “Dark Matter Searches at Colliders”. In: *Annual Review of Nuclear and Particle Science* 68.1 (2018), pages 429–459.

- [77] A. Albert et al. “Recommendations of the LHC Dark Matter Working Group: Comparing LHC searches for heavy mediators of dark matter production in visible and invisible decay channels”. In: (2017).
- [78] M. Milgrom. “A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis”. In: *The Astrophysical Journal* 270 (1983), pages 365–370. DOI: 10.1086/161130.
- [79] M. Milgrom. “A modification of the newtonian dynamics : implications for galaxy systems”. In: *The Astrophysical Journal* 270 (1983), pages 384–389. DOI: 10.1086/161132.
- [80] Y. J. Ko et al. “Comparison between DAMA/LIBRA and COSINE-100 in the light of Quenching Factors”. In: *Jcap* 11 (2019), page 008. DOI: 10.1088/1475-7516/2019/11/008.
- [81] J. H. Davis, C. McCabe, and C. Boehm. “Quantifying the evidence for dark matter in CoGeNT data”. In: *Journal of Cosmology and Astroparticle Physics* 2014.08 (2014), page 014. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2014/08/014.
- [82] P. Gondolo and G. Gelmini. “Cosmic abundances of stable particles: Improved analysis”. In: *Nuclear Physics B* 360.1 (1991), pages 145–179. ISSN: 0550-3213. DOI: 10.1016/0550-3213(91)90438-4.
- [83] D. Hooper and L. Goodenough. “Dark matter annihilation in the Galactic Center as seen by the Fermi Gamma Ray Space Telescope”. In: *Physics Letters B* 697.5 (2011), pages 412–428. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2011.02.029.
- [84] F. Calore, I. Cholis, and C. Weniger. “Background model systematics for the Fermi GeV excess”. In: *Journal of Cosmology and Astroparticle Physics* 2015.03 (2015), page 038. DOI: 10.1088/1475-7516/2015/03/038.
- [85] K. N. Abazajian and M. Kaplinghat. “Detection of a gamma-ray source in the Galactic Center consistent with extended emission from dark matter annihilation and concentrated astrophysical emission”. In: *Phys. Rev. D* 86.8 (2012), page 083511. DOI: 10.1103/PhysRevD.86.083511.
- [86] M. Aguilar et al. “First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–350 GeV”. In: *Phys. Rev. Lett.* 110.14 (2013), page 141102.
- [87] R. Cowsik, B. Burch, and T. Madziwa-Nussinov. “The Origin Of The Spectral Intensities Of Cosmic-Ray Positrons”. In: *The Astrophysical Journal* 786.2 (April 2014), page 124. DOI: 10.1088/0004-637x/786/2/124. URL: <https://doi.org/10.1088/0004-637x/786/2/124>.
- [88] K. Blum, B. Katz, and E. Waxman. “AMS-02 Results Support the Secondary Origin of Cosmic Ray Positrons”. In: *Phys. Rev. Lett.* 111.21 (2013), page 211101.

- [89] A. Reinert and M. W. Winkler. “A precision search for WIMPs with charged cosmic rays”. In: *Journal of Cosmology and Astroparticle Physics* 2018.01 (2018), page 055.
- [90] A. Coogan and S. Profumo. “Origin of the tentative AMS antihelium events”. In: *Phys. Rev. D* 96.8 (2017), page 083020.
- [91] J. W. Moffat. “Ultraviolet complete quantum field theory and particle model”. In: *The European Physical Journal Plus* 134.9 (2019), page 443. ISSN: 2190-5444. DOI: 10.1140/epjp/i2019-12973-6.
- [92] N. Trevisani. “Collider Searches for Dark Matter (ATLAS + CMS)”. In: *Universe* 4.11 (2018). ISSN: 2218-1997. DOI: 10.3390/universe4110131.
- [93] A. Boveia et al. “Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter”. In: *Physics of the Dark Universe* 27 (2020), page 100365. ISSN: 2212-6864. DOI: 10.1016/j.dark.2019.100365.
- [94] G. Aad et al. “A search for new physics in dijet mass and angular distributions in pp collisions at $\sqrt{s} = 7$ TeV measured with the ATLAS detector”. In: *New Journal of Physics* 13.5 (2011), page 053044. DOI: 10.1088/1367-2630/13/5/053044.
- [95] 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”. In: *Journal of High Energy Physics* 2019.5 (2019), page 142. ISSN: 1029-8479. DOI: 10.1007/Jhep05(2019)142.
- [96] J. L. A. Fernandez et al. “A Large Hadron Electron Collider at CERN Report on the Physics and Design Concepts for Machine and Detector”. In: *Journal of Physics G: Nuclear and Particle Physics* 39.7 (2012), page 075001. DOI: 10.1088/0954-3899/39/7/075001.
- [97] The ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (2008), page 08003. DOI: 10.1088/1748-0221/3/08/S08003.
- [98] CERN. *LHC Guide*. 2017.
- [99] J. T. Boyd. *LHC Run-2 and Future Prospects*. 2020.
- [100] M. Lamont. “Status of the LHC”. In: 455 (2013), page 012001. ISSN: 1742-6596. DOI: 10.1088/1742-6596/455/1/012001.
- [101] M. Bernardini and K. Foraz. “Long Shutdown 2 @ LHC”. In: (2015).
- [102] L. Evans and P. Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (2008), page 08001. DOI: 10.1088/1748-0221/3/08/S08001.
- [103] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: Cern, 2004. DOI: 10.5170/CERN-2004-003-V-1.

- [104] T. Lenzi. “Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors”. In: (2013).
- [105] C. Grupen and B. Shwartz. *Particle Detectors*. Cambridge University Press, 2008. ISBN: 9781139469531.
- [106] W. R. Leo. *Techniques for nuclear and particle physics experiments*. Springer Verlag, 1994.
- [107] J. Pequeno and P. Schaffner. *How ATLAS detects particles: diagram of particle paths in the detector*. 2013.
- [108] M. Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Technical report. 2010.
- [109] The ATLAS TRT Collaboration et al. “The ATLAS TRT Barrel Detector”. In: *Journal of Instrumentation* 3.02 (2008), page 02014. DOI: 10.1088/1748-0221/3/02/p02014.
- [110] G. Barrand et al. “GAUDI – A software architecture and framework for building HEP data processing applications”. In: *Computer Physics Communications* 140.1 (2001), pages 45–55. ISSN: 0010-4655. DOI: 10.1016/S0010-4655(01)00254-5.
- [111] S. Xella. *Physics objects reconstruction in the ATLAS experiment*. Technical report. Geneva, 2013.
- [112] L. Lönnblad. “CLHEP—a project for designing a C++ class library for high energy physics”. In: *Computer Physics Communications* 84.1 (1994), pages 307–316. ISSN: 0010-4655. DOI: 10.1016/0010-4655(94)90217-8.
- [113] T. Gleisberg et al. “Event generation with SHERPA 1.1”. In: *Journal of High Energy Physics* 2009.02 (2009), page 007.
- [114] T. Sjöstrand, S. Mrenna, and P. Skands. “A brief introduction to PYTHIA 8.1”. In: *Computer Physics Communications* 178.11 (2008), pages 852–867. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2008.01.036.
- [115] T. Sjöstrand, S. Mrenna, and P. Skands. “PYTHIA 6.4 physics and manual”. In: *Journal of High Energy Physics* 2006.05 (2006), page 026.
- [116] G. Corcella et al. “HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)”. In: *Journal of High Energy Physics* 2001.01 (2001), page 010.
- [117] G. Aad et al. “The ATLAS Simulation Infrastructure”. In: *The European Physical Journal C* 70.3 (2010), pages 823–874. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-010-1429-9.

- [118] S. Agostinelli et al. “Geant4 – a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pages 250–303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8.
- [119] J. Allison et al. “Geant4 developments and applications”. In: *IEEE Transactions on Nuclear Science* 53.1 (2006), pages 270–278. ISSN: 1558-1578.
- [120] J. Allison et al. “Recent developments in Geant4”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pages 186–225. ISSN: 0168-9002. DOI: 10.1016/j.nima.2016.06.125.
- [121] P. Calafiura et al. *ATLAS HL-LHC Computing Conceptual Design Report*. Technical report. Geneva, 2020.
- [122] R. Eckhardt. *Stan Ulam, John von Neumann, and the Monte Carlo method*. 1987. URL: http://www-star.st-and.ac.uk/~kw25/teaching/mcrt/MC_history_3.pdf.
- [123] D. Edwards. *Notes on Discrete Probability*. Unpublished. 2004.
- [124] *Introduction to Geant4*. URL: <https://indico.cern.ch/event/865808>.
- [125] *Geometry Description Markup Language (GDML)*. URL: <https://gdml.web.cern.ch/GDML/>.
- [126] *Geant4 Material Database*. URL: <https://geant4-userdoc.web.cern.ch/UsersGuides/ForApplicationDeveloper/html/Appendix/materialNames.html>.
- [127] *Geant4 Guide for Physics Lists*. URL: <https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsListGuide/html/index.html>.
- [128] *ROOT Data Analysis Framework*. URL: <https://root.cern.ch/>.
- [129] R. E. Bryant and D. R. O’Hallaron. *Computer Systems: A Programmer’s Perspective (2nd Edition)*. Prentice Hall, 2010. ISBN: 9780136108047.
- [130] C. Collberg et al. “Slinky: Static Linking Reloaded”. In: (2004).
- [131] A. Dotti. *Geant4 benchmark simulation*. 2018. URL: <https://gitlab.cern.ch/adotti/HepExpMT> (visited on February 3, 2021).
- [132] M. Matsumoto and T. Nishimura. “Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator”. In: *ACM Trans. Model. Comput. Simul.* 8.1 (1998), pages 3–30. ISSN: 1049-3301.
- [133] *CLHEP Library*. URL: <https://proj-clhep.web.cern.ch/proj-clhep/>.
- [134] *Geant4 Reference Physics Lists - Use cases*. URL: <https://geant4.web.cern.ch/node/302>.

- [135] V. V. Uzhinsky. “The Fritiof (FTF) Model in Geant4”. In: *International Conference on Calorimetry for the High Energy Frontier*. 2013.
- [136] A. Heikkinen, N. Stepanov, and J. P. Wellisch. “Bertini intranuclear cascade implementation in GEANT4”. In: *eConf C030324I* (2003), page 008.
- [137] *Geant4 General Particle Source*. URL: <https://geant4-userdoc.web.cern.ch/UsersGuides/ForApplicationDeveloper/html/GettingStarted/generalParticleSource.html>.
- [138] *Simple and Fast Physics List Description*. URL: https://www.slac.stanford.edu/comp/physics/geant4/slac_physics_lists/simple/physlistdoc.html.
- [139] M. Aaboud et al. “A measurement of the calorimeter response to single hadrons and determination of the jet energy scale uncertainty using LHC Run-1 pp-collision data with the ATLAS detector”. In: *The European Physical Journal C* 77.1 (2017), page 26. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-4580-0.
- [140] *The Geometry Description Markup Language*. URL: <https://gdml.web.cern.ch/GDML/>.
- [141] R. Chytráček et al. “Geometry Description Markup Language for Physics Simulation and Analysis Applications”. In: *IEEE Transactions on Nuclear Science* 53.5 (2006), pages 2892–2896. ISSN: 1558-1578. DOI: 10.1109/Tns.2006.881062.
- [142] J. Boudreau and V. Tsulaia. “The GeoModel Toolkit for Detector Description”. In: (2005).
- [143] M. Bandieramonte, J. Boudreau, and R. M. Bianchi. *FullSimLight: ATLAS standalone Geant4 simulation*. Technical report. Geneva, 2020.
- [144] C. Marcon and B. Morgan. *Building Geant4 as a single dynamic or static library*. <https://gitlab.cern.ch/atlas-simulation-team/geant4/-/commit/89704795358426a25d6518f5197762bdab19d007>. Accessed: 2021-02-25. 2020.
- [145] V. Agrawal et al. “Architectural Support for Dynamic Linking”. In: *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems*. Asplos’15. New York, NY, USA: Association for Computing Machinery, 2015, pages 691–702.
- [146] *Geant4 - Options for Changing the Compiler and Build Flag*. URL: <https://geant4-userdoc.web.cern.ch/UsersGuides/InstallationGuide/html/installguide.html#options-for-changing-the-compiler-and-build-flags>.
- [147] A. Lechner. “Particle interactions with matter”. In: *CERN Yellow Rep. School Proc.* 5 (2018). Edited by Bernhard Holzer. DOI: 10.23730/CYRSP-2018-005.47.

- [148] *Dates of GCC releases*. URL: <https://gcc.gnu.org/releases.html>.
- [149] *Changelog for GCC 5 Release Series*. URL: <https://gcc.gnu.org/gcc-5/changes.html>.
- [150] *Changelog for GCC 6 Release Series*. URL: <https://gcc.gnu.org/gcc-6/changes.html>.
- [151] *Changelog for GCC 7 Release Series*. URL: <https://gcc.gnu.org/gcc-7/changes.html>.
- [152] *Changelog for GCC 8 Release Series*. URL: <https://gcc.gnu.org/gcc-8/changes.html>.
- [153] G. Amadio et al. *GeantV: Results from the prototype of concurrent vector particle transport simulation in HEP*. 2020.
- [154] A. Dell’Acqua. Private communications. 2020.
- [155] M. Bandieramonte. Private communications. 2020.
- [156] *CERN Engineering Data Management Service (EDMS)*. URL: <https://edms-service.web.cern.ch/faq/EDMS/pages/>.
- [157] *CERN Drawing Directory (CDD)*. URL: <https://edms-service.web.cern.ch/CDD/>.
- [158] *ATLAS Detector Description Database (ADDD)*. URL: <https://atlas-geometry-db.web.cern.ch/atlas-geometry-db/>.
- [159] H. Ogren. Private communications. 2021.
- [160] *TRT Geometry Issue*. URL: <https://its.cern.ch/jira/browse/ATLASSIM-2217>.
- [161] C. Marcon. *Implementation of Arb8 shape for TRT modules*. 2020. URL: <https://gitlab.cern.ch/cmarcon/GeoModelPlugins/-/tree/test-make-module-with-arb8>.
- [162] C. Marcon. *Implementation of BRep shape for TRT modules*. 2020. URL: <https://gitlab.cern.ch/cmarcon/GeoModelPlugins/-/tree/test-make-module-with-brep>.
- [163] K. M. Zurek. “Asymmetric Dark Matter: Theories, signatures, and constraints”. In: *Physics Reports* 537.3 (2014), pages 91–121. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2013.12.001.
- [164] Y. Hochberg et al. “Mechanism for Thermal Relic Dark Matter of Strongly Interacting Massive Particles”. In: *Phys. Rev. Lett.* 113.17 (2014), page 171301.
- [165] I. Boucheneb et al. “Revisiting monotop production at the LHC”. In: *Journal of High Energy Physics* 2015.1 (2015), page 17. ISSN: 1029-8479. DOI: 10.1007/Jhep01(2015)017.

- [166] N. F. Bell, G. Busoni, and I. W. Sanderson. “Self-consistent Dark Matter simplified models with an s -channel scalar mediator”. In: *Journal of Cosmology and Astroparticle Physics* 2017.03 (2017), page 015.
- [167] M. Bauer, U. Haisch, and F. Kahlhoefer. “Simplified dark matter models with two Higgs doublets: I. Pseudoscalar mediators”. In: *Journal of High Energy Physics* 2017.5 (2017), page 138. ISSN: 1029-8479. DOI: 10.1007/Jhep05(2017)138.
- [168] M. Duerr et al. “How to save the WIMP: global analysis of a dark matter model with two s -channel mediators”. In: *Journal of High Energy Physics* 2016.9 (2016), page 42. ISSN: 1029-8479. DOI: 10.1007/Jhep09(2016)042.
- [169] J. Goodman et al. “Constraints on dark matter from colliders”. In: *Phys. Rev. D* 82.11 (2010), page 116010.
- [170] C. D. Carone and H. Murayama. “Possible Light $U(1)$ Gauge Boson Coupled to Baryon Number”. In: *Phys. Rev. Lett.* 74.16 (1995), pages 3122–3125.
- [171] L. Carpenter et al. “Mono-Higgs-boson: A new collider probe of dark matter”. In: *Phys. Rev. D* 89.7 (2014), page 075017.
- [172] J. Andrea, B. Fuks, and F. Maltoni. “Monotops at the LHC”. In: *Phys. Rev. D* 84.7 (2011), page 074025.
- [173] P. J. Fox et al. “Missing energy signatures of dark matter at the LHC”. In: *Phys. Rev. D* 85.5 (2012), page 056011.
- [174] P. Gondolo, P. Ko, and Y. Omura. “Light dark matter in leptophobic Z' models”. In: *Phys. Rev. D* 85.3 (2012), page 035022.
- [175] O. Buchmueller et al. “Characterising dark matter searches at colliders and direct detection experiments: vector mediators”. In: *Journal of High Energy Physics* 2015.1 (2015), page 37. ISSN: 1029-8479. DOI: 10.1007/Jhep01(2015)037.
- [176] M. Aaboud et al. “Search for new high-mass phenomena in the dilepton final state using 36 fb⁻¹ of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2017.10 (2017), page 182. ISSN: 1029-8479. DOI: 10.1007/Jhep10(2017)182.
- [177] V. Khachatryan et al. “Search for narrow resonances in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 13$ TeV and combination with 8 TeV data”. In: *Physics Letters B* 768 (2017), pages 57–80. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2017.02.010.
- [178] A. M. Sirunyan et al. “Search for new physics in dijet angular distributions using proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter and other models”. In: *The European Physical Journal C* 78.9 (2018), page 789. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-6242-x.

- [179] M. Aaboud et al. “Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector”. In: *Phys. Rev. D* 96.5 (2017), page 052004.
- [180] G. Arnison et al. “Angular distributions and structure functions from two-jet events at the CERN SPS collider”. In: *Physics Letters B* 136.4 (1984), pages 294–300. ISSN: 0370-2693. DOI: 10.1016/0370-2693(84)91164-X.
- [181] V. M. Abazov et al. “Measurement of Dijet Angular Distributions at $\sqrt{s} = 1.96 \text{ TeV}$ and Searches for Quark Compositeness and Extra Spatial Dimensions”. In: *Phys. Rev. Lett.* 103.19 (2009), page 191803.
- [182] The ATLAS Collaboration. “Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector”. In: 788 (2019), page 316. DOI: 10.1016/j.physletb.2018.09.062.
- [183] A. M. Sirunyan et al. “Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ”. In: *Journal of High Energy Physics* 2018.1 (2018), page 97. ISSN: 1029-8479. DOI: 10.1007/Jhep01(2018)097.
- [184] The ATLAS Collaboration. *Search for light dijet resonances with the ATLAS detector using a Trigger-Level Analysis in LHC pp collisions at $\sqrt{s} = 13 \text{ TeV}$* . Technical report. Geneva, 2016.
- [185] The ATLAS Collaboration. *Trigger-object Level Analysis with the ATLAS detector at the Large Hadron Collider: summary and perspectives*. Technical report. Geneva, 2017.
- [186] R. Aaij et al. “Tesla: An application for real-time data analysis in High Energy Physics”. In: *Computer Physics Communications* 208 (2016), pages 35–42. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2016.07.022.
- [187] V. Khachatryan et al. “Search for Narrow Resonances in Dijet Final States at $\sqrt{s} = 8 \text{ TeV}$ with the Novel CMS Technique of Data Scouting”. In: *Physical Review Letters* 117.3 (2016), page 031802. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.117.031802.
- [188] D. Abercrombie et al. “Dark Matter benchmark models for early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum”. In: *Physics of the Dark Universe* 27 (2020), page 100371. ISSN: 2212-6864. DOI: 10.1016/j.dark.2019.100371.
- [189] S. Schramm et al. “ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz-boosted Objects”. In: *EPJ Web of Conferences* 182 (2018), page 02113. ISSN: 2100-014X. DOI: 10.1051/epjconf/201818202113.

- [190] M. Aaboud et al. “Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2”. In: *The European Physical Journal C* 77.10 (2017), page 673. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5225-7.
- [191] J. J. Goodson. “Search for Supersymmetry in States with Large Missing Transverse Momentum and Three Leptons including a Z-Boson”. In: (2012).
- [192] G. Aad et al. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *The European Physical Journal C* 77.7 (2017), page 490. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5004-5.
- [193] S. D. Ellis and D. E. Soper. “Successive combination jet algorithm for hadron collisions”. In: *Phys. Rev. D* 48.7 (1993), pages 3160–3166.
- [194] R. Akhoury, M. G. Sotiropoulos, and V. I. Zakharov. “Kinoshita-Lee-Nauenberg theorem and soft radiation in gauge theories: Abelian case”. In: *Phys. Rev. D* 56.1 (1997), pages 377–387.
- [195] Y. L. Dokshitzer et al. “Better jet clustering algorithms”. In: 1997.08 (1997), page 001. ISSN: 1029-8479. DOI: 10.1088/1126-6708/1997/08/001.
- [196] M. Cacciari, G. P. Salam, and G. Soyez. “The anti-kt jet clustering algorithm”. In: 2008.04 (2008), page 063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063.
- [197] G. Aad et al. “Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *The European Physical Journal C* 81.8 (2021), page 689. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09402-3.
- [198] M. Aaboud et al. “In situ calibration of large-radius jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector”. In: *The European Physical Journal C* 79.2 (2019). ISSN: 1434-6044. DOI: 10.1140/epjc/s10052-019-6632-8.
- [199] M. Aaboud et al. “Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Physical Review D* 96.7 (2017). ISSN: 2470-0010. DOI: 10.1103/PhysRevD.96.072002.
- [200] G. Cowan. “Statistics for Searches at the LHC”. In: (2013).
- [201] K. Pearson. “X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50.302 (1900), pages 157–175. ISSN: 1941-5982. DOI: 10.1080/14786440009463897.
- [202] I. Narsky and F. C. Porter. *Statistical Analysis Techniques in Particle Physics*. Wiley-VCH Verlag GmbH & Co. KGaA, 2013.
- [203] G. Choudalakis. “On hypothesis testing, trials factor, hypertests and the BumpHunter”. In: (2018).

- [204] O. Behnke et al., editors. *Data Analysis in High Energy Physics*. Wiley-VCH Verlag GmbH & Co. KGaA, 2013.
- [205] 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”. In: *Physical Review Letters* 121.8 (2018), page 081801. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.121.081801.
- [206] G. Aad et al. “Luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC”. In: *The European Physical Journal C* 71.4 (2011), page 1630. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1630-5.
- [207] G. Aad et al. “Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC”. In: *The European Physical Journal C* 73.8 (2013), page 2518. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2518-3.
- [208] C. Röver, C. Messenger, and R. Prix. “Bayesian versus frequentist upper limits”. In: *Proceedings of the PHYSTAT 2011* (2011), pages 17–20.
- [209] S. Mukherjee. *Data Scouting: A New Trigger Paradigm*. 2017.
- [210] The ATLAS Collaboration. “Operation of the ATLAS trigger system in Run 2”. In: *Journal of Instrumentation* 15.10 (2020), P10004–P10004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/10/P10004.
- [211] The ATLAS Collaboration. “Performance of the missing transverse momentum triggers for the ATLAS detector during Run-2 data taking”. In: *Jhep* 08 (2020), page 80. DOI: 10.1007/JHEP08(2020)080.
- [212] The ATLAS Collaboration. *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*. Technical report. Geneva, 2015.
- [213] M. Cacciari and G. P. Salam. “Dispelling the N₃ myth for the kt jet-finder”. In: *Physics Letters B* 641.1 (2006), pages 57–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2006.08.037.
- [214] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Jhep* 07 (2014), page 079. DOI: 10.1007/JHEP07(2014)079.
- [215] T. Sjöstrand, S. Mrenna, and P. Skands. “A Brief Introduction to PYTHIA 8.1”. In: 178 (2008), pages 852–867. DOI: 10.1016/j.cpc.2008.01.036.
- [216] *BayesianFramework source code*. URL: https://gitlab.cern.ch/atlas-phys-exotics-dijet-tla/BayesianFramework_FullRun2.
- [217] *Limit Setting and Spurious Signal in the Dijet Trigger Level Analysis*. URL: https://indico.cern.ch/event/913589/contributions/3841849/attachments/2034307/3405608/DijetTLA_StatsComSlides.pdf.

- [218] *FrequentistFramework (HistFitter) source code*. URL: <https://gitlab.cern.ch/atlas-phys-exotics-dijet-tla/FrequentistFramework>.
- [219] *XmlAnaWSBuilder source code*. URL: <https://gitlab.cern.ch/atlas-hgam-sw/xmlAnaWSBuilder>.
- [220] *QuickFit source code*. URL: <https://gitlab.cern.ch/atlas-phys-exotics-dijet-tla/quickFit>.
- [221] N. Berger et al. *Recommendations for the Modeling of Smooth Backgrounds*. Technical report. Geneva, 2020.
- [222] D.F. Morrison et al. *Multivariate Statistical Methods*. Annals of the New York Academy of Sciences. McGraw-Hill, 1976. ISBN: 9780070431867.
- [223] *ROOT-based Prefit Tool for TLA background fit*. URL: <https://cernbox.cern.ch/index.php/s/W0n2jk3j9d7nbZW>.

