

ABSTRACT

RNTUPLE FOR ATLAS ANALYSIS WORKFLOWS

Fatima Rodriguez, M.A.
Department of Physics
Northern Illinois University, 2025
Hector de la Torre, Director

RNTuple is the new data storage format set to replace TTree at the start of the High Luminosity LHC. An investigation was conducted on how analysis workflows for ATLAS researchers will change with RNTuple. Additionally, performance studies have been conducted that demonstrate an improvement in speed and memory usage at the analysis front. Finally, different compression algorithms were tested and it was found that blah blah remains to best work with RNTuple.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

DECEMBER 2025

RNTUPLE FOR ATLAS ANALYSIS WORKFLOWS

BY

FATIMA RODRIGUEZ
© 2025 Fatima Rodriguez

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
MASTER OF PHYSICS

DEPARTMENT OF PHYSICS

Dissertation Director:
Hector de la Torre

13

ACKNOWLEDGEMENTS

14

Thanks thanks

DEDICATION

To my mum.

TABLE OF CONTENTS

Chapter	Page
¹⁷ 1 Introduction	¹
¹⁸ 1.0.1 Standard model of particle physics	³
¹⁹ 1.0.1.1 Symmetries and Particle Content	³
²⁰ 1.0.2 Standard Model Limitations	⁵
²¹ 1.0.3 Phenomenology of Large Hadron Colliders	⁵
²² 2 The ATLAS Experiment	⁶
²³ 2.1 Detector Technology	⁷
²⁴ 2.1.1 Inner Detector	⁷
²⁵ 2.1.1.1 Pixel Detector	⁹
²⁶ 2.1.1.2 Semiconductor Tracker	⁹
²⁷ 2.1.1.3 Transition Radiation Tracker	⁹
²⁸ 2.1.2 Calorimeter	⁹
²⁹ 2.1.2.1 Liquid Argon Calorimeter	⁹
³⁰ 2.1.3 Tile Hadronic Calorimeter	⁹
³¹ 2.1.4 Muon Spectrometer	⁹
³² 2.1.4.1 Thin Gap Chambers	⁹
³³ 2.1.4.2 Resistive Plate Chambers	⁹
³⁴ 2.1.4.3 Monitored Drift Tubes	⁹
³⁵ 2.1.4.4 Small-Strip Thin-Gap	⁹
³⁶ 2.1.4.5 Micromegas	⁹

	Chapter		Page
37			
38	2.1.5	Magnet System	9
39	2.1.5.1	Central Solenoid Magnet	9
40	2.1.5.2	Barrel Toroid	9
41	2.1.5.3	End-cap Toroids	9
42	2.2	ATLAS Trigger System	9
43	2.2.1	First-Level Hardware Trigger	9
44	2.2.2	Second-Level Hardware Trigger	9
45	References		10
46	Appendix: Objective Symptoms		11

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

51 Our current understanding of the building blocks of our universe is summarized with one
52 model, called the Standard Model (SM). From the way we power our cities, to the particles
53 that hold them together, the SM explains how the basic building blocks of matter interact,
54 governed by the four fundamental forces: gravity, electromagnetism, the strong force and the
55 weak force. Yet, questions remain about the SM, such as why are there only three generations
56 of fundamental particles? What is the nature of dark matter and dark energy, and how does
57 it fit within the SM? What about the origin of the matter-antimatter asymmetry? Is there a
58 unification theory for the fundamental forces? Is the SM complete or do other exotic particles
59 exists? Over the years, experimental particle physicists and engineers have built technology
60 to test the SM, either by performing precision measurements of particles and their behaviors,
61 or by colliding particles and measuring their outputs. As a result, we have increased our
62 confidence in the SM theory, but continue to search for answers for these remaining questions
63 through experimental discovery.

64 A Toroidal LHC Apparatus (ATLAS) is a particle physics experiment designed to detect
65 the high-energy particle collisions from the Large Hadron Collider (LHC). At the LHC,
66 collisions take place at a rate of more than a billion interactions per second, which is a
67 combined data volume of about 60 million megabytes per second. However, in order to
68 study rare processes, as shown in Figure 1.1, the LHC will have a major upgrade to increase
69 the number of collisions by a factor of 5 to 7.5. This upgrade, called the High-Luminosity
70 LHC, will require a new data storage format that can handle this increase in data.

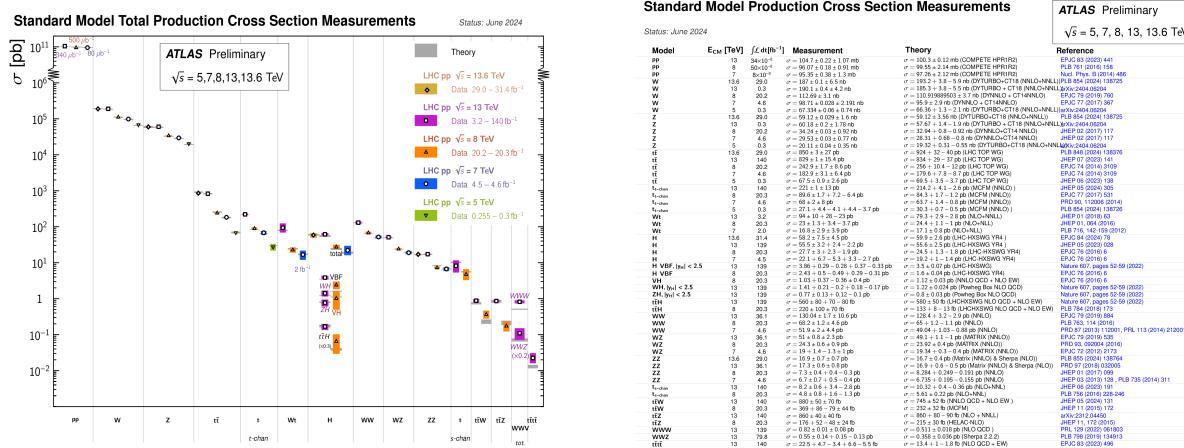


Figure 1.1: Summary of several Standard Model cross-section measurements (a) with associated references (b) [1]. Cross-sections can be thought of as probabilities that a process occurs. This means that the processes with smaller cross-sections are considered rare-processes because it has a lower probability of being observed. Increasing the probability of these rare-processes would require an increase of luminosity or collisions. The measurements are corrected for branching fractions, compared to the corresponding theoretical expectations.

RNTuple is the new ROOT data storage format that will be in use at the start of the HL-LHC. Due to its design, which takes advantage of modern C++ techniques, it is set to improve read speedability and memory usage compared to its predecessor, TTree, and other data storage formats such as HDF5 and Parquet. At the start of this work, performance studies on RNTuple were conducted at the production level, and RNTuple was still at an experimental stage.

This thesis investigates the performance of RNTuple for ATLAS analysis workflows. This chapter will provide a more detail introduction of the SM, followed by an introduction to the ATLAS experiment and its detector technology in Chapter 2. In Chapter 3, the ATLAS software and computing system, and data contents are introduced. In Chapter 4, an introduction to RNTuple and TTree is provided along with examples of how RNTuple is applied in comparison to TTree. Performance studies conducted for RNTuple and how they compare with TTree will be presented in Chapter 5. In Chapter 6, the Analysis Grand

⁸⁴ Challenge (AGC) is introduced along with its RNTuple implementation. A final discussion
⁸⁵ and conclusions are given in Chapter 7.

⁸⁶ **1.0.1 Standard model of particle physics**

⁸⁷ The SM is a quantum field theory that explains and catagorizes all observed fundamental
⁸⁸ particles by their properties and interactions. Quantum field theory (QFT) is the main the-
⁸⁹ oretical tool for describing particle interactions by combining special relativity and quantum
⁹⁰ mechanics. Due to this combination, QFT is a probabilistic theory where each particle has
⁹¹ an associated field that permeates all of space; therefore, forces are simply the interactions
⁹² between these different fields. For example, the electromagnetic force is just the interac-
⁹³ tion between the electromagnetic field and charged matter fields, which fall under quantum
⁹⁴ electrodynamics (QED). In sum, the SM encompasses all known elementary particle inter-
⁹⁵ actions, except for gravity, through a collection of quantum field theories, each dictated by
⁹⁶ gauge symmetries: QED ($U(1)$), the Glashow-Weinberg-Salam theory of electroweak pro-
⁹⁷ cesses ($SU(3)$), and quantum chromodynamics ($SU(2)xU(1)$).

⁹⁸ **1.0.1.1 Symmetries and Particle Content**

⁹⁹ In physics, symmetries are fundamental because they lead to conservation laws through
¹⁰⁰ Noether's Theorem. Symmetries can manifest in two notions: invariance and covariance.
¹⁰¹ Properties of a system are described as invariant if they do not change under a symme-
¹⁰² try transformation. For example, rotating a sphere and without altering gravitational force
¹⁰³ would indicate a conservation of angular momentum. In contrast, covariance is used to de-
¹⁰⁴ scribe a system that changes in accordance to changes induced by symmetry transformations.

105 The SM is a gauge theory based on the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. Gauge
 106 theory is a QFT that requires invariance under continuous transformations, and a symmetry
 107 group is a set objects that obey the four properties listed in Table 1.1. $SU(3)_c$ is called the
 108 color symmetry group describing the strong nuclear force, which is the interaction between
 109 quarks and gluons. $SU(2)_L \times U(1)_Y$ describes the electromagnetic and weak nuclear forces,
 which is the interactions between leptons, photons and WZ bosons.

Table 1.1: Properties of a Group [4].

CLOSURE	If $g_1, g_2 \in G \rightarrow g_1 \diamond g_2 \in G$	(combinations remain in G)
IDENTITY	There exists $I \in G \rightarrow I \diamond g_i = g_i$ for every $g_i \in G$	(one element does not change others)
INVERSE	Every $g_i \in G$ has a $g_i^{-1} \in G$ such that $g_1 \diamond g_i^{-1} = I$	(combinations can be undone)
ASSOCIATIVITY	If $g_1, g_2, g_3 \in G \rightarrow (g_1 \diamond g_2) \diamond g_3 = g_1 \diamond (g_2 \diamond g_3)$	(combinational groupings can be rearranged)

110
 111 Furthermore, the four groups of particles shown in Figure 1.2: quarks, leptons, gauge
 112 bosons, and scalar bosons, can be further categorized as *bosons* or *fermions* because of
 113 a fundamental property called spin. Similar to the Earth, particles carry orbital angular
 114 momentum and spin angular momentum; however, for particles, spin is an intrinsic property.
 115 All bosons carry an integer spin; meanwhile, fermions carry half-integer spin. As a result
 116 from QFT, each fermion has an antiparticle with the same mass and lifetime as the particle
 117 itself, but oppositely charged. The three charged leptons (e, μ, τ) are massive, while their
 118 corresponding neutrinos (ν_e, ν_μ, ν_τ), are massless with neutral charge. Due to QCD, there are
 119 8 types of gluons. The Higgs boson has its own section as a scalar boson because unlike the
 120 vector bosons with spin 1, the Higgs boson has spin 0. In sum, there are a total of 12 leptons
 121 including their antiparticles, 36 quarks including all the flavors and their antiparticles, 12
 122 vector bosons, and 1 scalar boson, which makes a total of 61 fundamental particles.

Standard Model of Elementary Particles

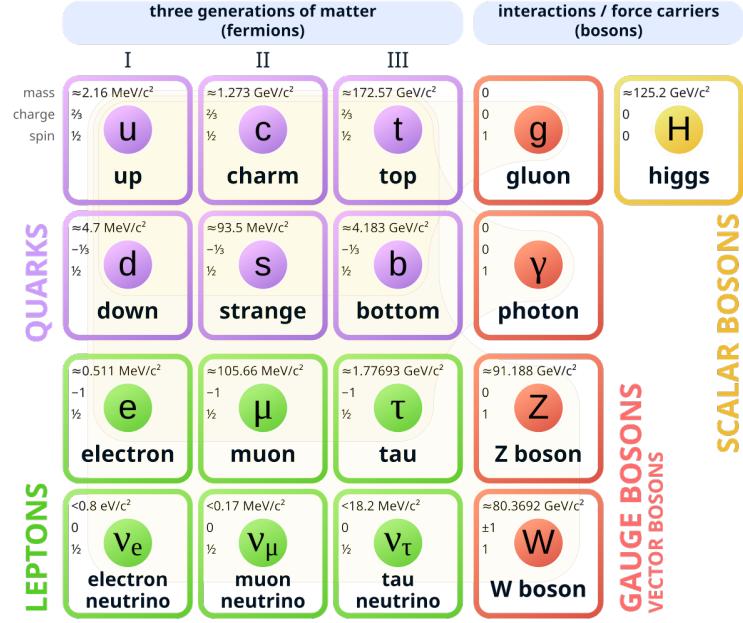


Figure 1.2: Particle content of the Standard Model [?].

123

1.0.2 Standard Model Limitations

124

1.0.3 Phenomenology of Large Hadron Colliders

125

CHAPTER 2

126

THE ATLAS EXPERIMENT

127 In October 1992, the ATLAS collaboration, then composed of about 800 members, submitted a letter of intent to the LHC Experiment Committee highlighting the design of what 128 came to be today's ATLAS Experiment ??. From the start, ATLAS was designed to be 129 a general-purpose experiment, optimized to search for the Higgs boson, top quark decays, 130 and supersymmetry. In July 1997, the ATLAS Experiment was approved and by November 131 2008, ATLAS was the largest detector ever constructed at 44 meters long and 25 meters 132 in diameter. By November 2009, ATLAS recorded its first proton-proton collision and by 133 December 2010, ATLAS was first to observed the production of top quark pairs, which are 134 the heaviest known elementary particle with a strong coupling to the Higgs boson. By July 135 2012, both ATLAS and the CMS Experiment successfully observed the infamous Higgs 136 boson. ATLAS is projected to continue operation until 2035 to continue searching for standing 137 questions from the SM.

139 In par with other particle experiments, the ATLAS detector has two general components: 140 calorimeters and magnets. Calorimteter

141 : calorimeters and magnets. Calorimeters are devices that measures the energy a particle 142 loses when passing through. Through

143

2.1 Detector Technology

144 To do this, ATLAS has six different detecting subsystems wrapped concentrically in
145 layers around the collision point to record the trajectory, momentum, and energy of particles.

146 Apart, a huge magnet system bends the paths of the charged particles so that their momenta
147 can be measured as precisely as possible. Overall, the detector tracks and identifies particles
148 to investigate a wide range of physics.

149

2.1.1 Inner Detector

150 what is its main functions

¹⁵¹ 2.1.1.1 Pixel Detector

¹⁵² 2.1.1.2 Semiconductor Tracker

¹⁵³ 2.1.1.3 Transition Radiation Tracker

¹⁵⁴ **2.1.2 Calorimeter**

¹⁵⁵ 2.1.2.1 Liquid Argon Calorimeter

¹⁵⁶ 2.1.3 Tile Hadronic Calorimeter

¹⁵⁷ **2.1.4 Muon Spectrometer**

¹⁵⁸ 2.1.4.1 Thin Gap Chambers

¹⁵⁹ 2.1.4.2 Resistive Plate Chambers

¹⁶⁰ 2.1.4.3 Monitored Drift Tubes

¹⁶¹ 2.1.4.4 Small-Strip Thin-Gap

¹⁶² 2.1.4.5 Micromegas

¹⁶³ **2.1.5 Magnet System**

REFERENCES

- ¹⁷¹ [1] ATLAS Collaboration. Standard Model summary plots. *ATLAS Public Note ATL-*
¹⁷² *PHYS-PUB-2024-011*, 2024.
- ¹⁷³ [2] ATLAS Collaboration ATLAS: Letter of intent for a general purpose p p experiment at
¹⁷⁴ the large hadron collider at CERN. *CERN-LHCC-92-04*, 1992.
- ¹⁷⁵ [3] ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider:
¹⁷⁶ a description of the detector configuration for Run 3. *JINST*, 19:P05063, 2024. URL:
¹⁷⁷ <http://dx.doi.org/10.1088/1748-0221/19/05/P05063>
- ¹⁷⁸ [4] R. Mann. *An Introduction to Particle Physics and the Standard Model*. Taylor & Francis,
¹⁷⁹ 2010. ISBN 978-1-4200-8300-2.
- ¹⁸⁰ [5] Wikipedia. Elementary particle. URL: https://en.wikipedia.org/wiki/Elementary_particle. Accessed: October 9, 2025.

182

APPENDIX

183

OBJECTIVE SYMPTOMS

¹⁸⁴ Appendices follow the same page-numbering rules as regular chapters. The first page of a
¹⁸⁵ multi-page appendix is not numbered. But the page of a single-page appendix *is* numbered.

¹⁸⁶ **Are they slow learners** or is it a *REAL* problem? These are classic findings in the
¹⁸⁷ hopelessly computer challenged.

¹⁸⁸ 1. Can't copy from hard drive to disk.

¹⁸⁹ 2. Can't eject disks.

¹⁹⁰ 3. The word "disk" has thousands of meanings to them. None are correct.

¹⁹¹ 4. Saving a document in any form is a concept totally unexplainable to them.

¹⁹² 5. Desktop covered with Untitled Folders - look again, untitled folders are everywhere.

¹⁹³ 6. "Lost" documents found often in the Apple Menu.

¹⁹⁴ 7. Trash always full. Claim they don't know how to place things in trash.

¹⁹⁵ 8. Mysterious things happen to their documents or computer when they are not present.

¹⁹⁶ AKA "computer victims".

¹⁹⁷ 9. Highlighting = deleting. Dragging = Oblivion.

¹⁹⁸ 10. Selecting, double-clicking a problem? They will always say their mouse is broken.

¹⁹⁹ 11. Their double- click mechanics wants you to send them to a neurologist.

²⁰⁰ 12. Computer always on due to fear of having to restart it.

²⁰¹ 13. Have never read their QuickMail - will say "I prefer a phone call".

²⁰² 14. Have magical beliefs about what computers do.

²⁰³ 15. Describes some flaky way computers could REALLY help them, but is not yet available.

- ²⁰⁴ 16. Constantly saying they need more “memory”.
- ²⁰⁵ 17. Requests gizmos and gadgets, i.e., “mouse leash” or “disk cozy”.
- ²⁰⁶ 18. Avoids eye contact when talking about computers.