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

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

ACKNOWLEDGEMENTS

Thanks thanks

DEDICATION

To my mum.

TABLE OF CONTENTS

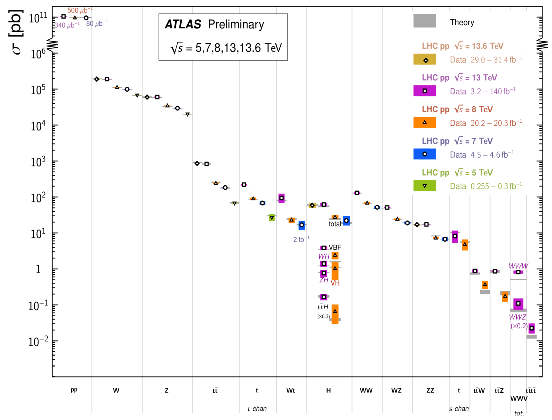
CHAPTER 1

INTRODUCTION

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

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

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



Standard Model Production Cross Section Measurements

Status: June 2024

ATLAS Preliminary
$\sqrt{s}$ = 5, 7, 8, 13, 13.6 TeV

Model	E_{CM} [TeV]	$\int \mathcal{L}\,dt$ [fb^{-1}]	Measurement	Theory	Reference
pp	13	34×10^{-8}	$\sigma = 104.7 \pm 0.22 \pm 1.07$ mb	$\sigma = 100.3 \pm 0.12$ mb (COMPETE HPR1R2)	EPJC 83 (2023) 441
pp	8	50×10^{-8}	$\sigma = 96.07 \pm 0.18 \pm 0.91$ mb	$\sigma = 99.55 \pm 2.14$ mb (COMPETE HPR1R2)	PLB 761 (2016) 158
pp	7	8×10^{-8}	$\sigma = 95.35 \pm 0.38 \pm 1.3$ mb	$\sigma = 97.26 \pm 2.12$ mb (COMPETE HPR1R2)	Nucl. Phys. B (2014) 486
W	13.6	29.0	$\sigma = 187 \pm 0.1 \pm 6.5$ nb	$\sigma = 193.2 + 3.8 - 5.9$ nb (DYTURBO+CT18 (NNLO+NNLL))	PLB 854 (2024) 138725
W	13	0.3	$\sigma = 190.1 \pm 0.4 \pm 4.2$ nb	$\sigma = 185.3 + 3.8 - 5.5$ nb (DYTURBO + CT18 (NNLO+NNLL))	arXiv:2404.06204
W	8	20.2	$\sigma = 112.69 \pm 3.1$ nb	$\sigma = 110.919889503 \pm 3.7$ nb (DYNNLO + CT14NNLO)	EPJC 79 (2019) 760
W	7	4.6	$\sigma = 98.71 \pm 0.028 \pm 2.191$ nb	$\sigma = 95.9 \pm 2.9$ nb (DYNNLO + CT14NNLO)	EPJC 77 (2017) 367
W	5	0.3	$\sigma = 67.334 \pm 0.06 \pm 0.74$ nb	$\sigma = 66.36 + 1.3 - 2.1$ nb (DYTURBO+CT18 (NNLO+NNLL))	arXiv:2404.06204
Z	13.6	29.0	$\sigma = 59.12 \pm 0.029 \pm 1.6$ nb	$\sigma = 59.12 \pm 3.56$ nb (DYTURBO+CT18 (NNLO+NNLL))	PLB 854 (2024) 138725
Z	13	0.3	$\sigma = 60.18 \pm 0.2 \pm 1.78$ nb	$\sigma = 57.67 + 1.4 - 1.9$ nb (DYTURBO + CT18 (NNLO+NNLL))	arXiv:2404.06204
Z	8	20.2	$\sigma = 34.24 \pm 0.03 \pm 0.92$ nb	$\sigma = 32.94 + 0.8 - 0.92$ nb (DYNNLO+CT14 NNLO)	JHEP 02 (2017) 117
Z	7	4.6	$\sigma = 29.53 \pm 0.03 \pm 0.77$ nb	$\sigma = 28.31 + 0.68 - 0.8$ nb (DYNNLO+CT14 NNLO)	JHEP 02 (2017) 117
Z	5	0.3	$\sigma = 20.11 \pm 0.04 \pm 0.35$ nb	$\sigma = 19.32 + 0.31 - 0.55$ nb (DYTURBO+CT18 (NNLO+NNLL))	arXiv:2404.06204
$t\bar{t}$	13.6	29.0	$\sigma = 850 \pm 3 \pm 27$ pb	$\sigma = 924 + 32 - 40$ pb (LHC TOP WG)	PLB 848 (2024) 138376
$t\bar{t}$	13	140	$\sigma = 829 \pm 1 \pm 15.4$ pb	$\sigma = 834 + 29 - 37$ pb (LHC TOP WG)	JHEP 07 (2023) 141
$t\bar{t}$	8	20.2	$\sigma = 242.9 \pm 1.7 \pm 8.6$ pb	$\sigma = 256 + 10.4 - 12$ pb (LHC TOP WG)	EPJC 74 (2014) 3109
$t\bar{t}$	7	4.6	$\sigma = 182.9 \pm 3.1 \pm 6.4$ pb	$\sigma = 179.6 + 7.8 - 8.7$ pb (LHC TOP WG)	EPJC 74 (2014) 3109
$t\bar{t}$	5	0.3	$\sigma = 67.5 \pm 0.9 \pm 2.6$ pb	$\sigma = 69.5 + 3.5 - 3.7$ pb (LHC TOP WG)	JHEP 06 (2023) 138
t_{t-chan}	13	140	$\sigma = 221 \pm 1 \pm 13$ pb	$\sigma = 214.2 + 4.1 - 2.6$ pb (MCFM (NNLO))	JHEP 05 (2024) 305
t_{t-chan}	8	20.3	$\sigma = 89.6 \pm 1.7 + 7.2 - 6.4$ pb	$\sigma = 84.3 + 1.7 - 1.2$ pb (MCFM (NNLO))	EPJC 77 (2017) 531
t_{t-chan}	7	4.6	$\sigma = 68 \pm 2 \pm 8$ pb	$\sigma = 63.7 + 1.4 - 0.8$ pb (MCFM (NNLO))	PRD 90, 112006 (2014)
t_{t-chan}	5	0.3	$\sigma = 27.1 + 4.4 - 4.1 + 4.4 - 3.7$ pb	$\sigma = 30.3 + 0.7 - 0.5$ pb (MCFM (NNLO))	PLB 854 (2024) 138726
Wt	13	3.2	$\sigma = 94 \pm 10 + 28 - 23$ pb	$\sigma = 79.3 + 2.9 - 2.8$ pb (NLO+NNLL)	JHEP 01 (2018) 63
Wt	8	20.3	$\sigma = 23 \pm 1.3 + 3.4 - 3.7$ pb	$\sigma = 24.4 + 1.1 - 1$ pb (NLO+NLL)	JHEP 01, 064 (2016)
Wt	7	2.0	$\sigma = 16.8 \pm 2.9 \pm 3.9$ pb	$\sigma = 17.1 \pm 0.8$ pb (NLO+NLL)	PLB 716, 142-159 (2012)
H	13.6	31.4	$\sigma = 58.2 \pm 7.5 \pm 4.5$ pb	$\sigma = 59.9 \pm 2.6$ pb (LHC-HXSWG YR4)	EPJC 84 (2024) 78
H	13	139	$\sigma = 55.5 \pm 3.2 + 2.4 - 2.2$ pb	$\sigma = 55.6 \pm 2.5$ pb (LHC-HXSWG YR4)	JHEP 05 (2023) 028
H	8	20.3	$\sigma = 27.7 \pm 3 + 2.3 - 1.9$ pb	$\sigma = 24.5 + 1.3 - 1.8$ pb (LHC-HXSWG YR4)	EPJC 76 (2016) 6
H	7	4.5	$\sigma = 22.1 + 6.7 - 5.3 + 3.3 - 2.7$ pb	$\sigma = 19.2 + 1 - 1.4$ pb (LHC-HXSWG YR4)	EPJC 76 (2016) 6
H VBF, $\|y_H\| < 2.5$	13	139	$\sigma = 3.86 + 0.29 - 0.28 + 0.37 - 0.33$ pb	$\sigma = 3.5 \pm 0.07$ pb (LHC-HXSWG)	Nature 607, pages 52-59 (2022)
H VBF	8	20.3	$\sigma = 2.43 + 0.5 - 0.49 + 0.29 - 0.31$ pb	$\sigma = 1.6 \pm 0.04$ pb (LHC-HXSWG YR4)	EPJC 76 (2016) 6
VH	8	20.3	$\sigma = 1.03 + 0.37 - 0.36 \pm 0.4$ pb	$\sigma = 1.12 \pm 0.03$ pb (NNLO QCD + NLO EW)	EPJC 76 (2016) 6
WH, $\|y_H\| < 2.5$	13	139	$\sigma = 1.41 + 0.21 - 0.2 + 0.18 - 0.17$ pb	$\sigma = 1.22 \pm 0.024$ pb (Powheg Box NLO QCD)	Nature 607, pages 52-59 (2022)
ZH, $\|y_H\| < 2.5$	13	139	$\sigma = 0.77 \pm 0.13 + 0.12 - 0.1$ pb	$\sigma = 0.8 \pm 0.03$ pb (Powheg Box NLO QCD)	Nature 607, pages 52-59 (2022)
$t\bar{t}H$	13	139	$\sigma = 560 \pm 80 + 70 - 80$ fb	$\sigma = 580 \pm 50$ fb (LHCHXSWG NLO QCD + NLO EW)	Nature 607, pages 52-59 (2022)
$t\bar{t}H$	8	20.3	$\sigma = 220 \pm 100 \pm 70$ fb	$\sigma = 133 + 8 - 13$ fb (LHCHXSWG NLO QCD + NLO EW)	PLB 784 (2018) 173
WW	13	36.1	$\sigma = 130.04 \pm 1.7 \pm 10.6$ pb	$\sigma = 128.4 + 3.2 - 2.9$ pb (NNLO)	EPJC 79 (2019) 884
WW	8	20.3	$\sigma = 68.2 \pm 1.2 \pm 4.6$ pb	$\sigma = 65 + 1.2 - 1.1$ pb (NNLO)	PLB 763, 114 (2016)
WW	7	4.6	$\sigma = 51.9 \pm 2 \pm 4.4$ pb	$\sigma = 49.04 + 1.03 - 0.88$ pb (NNLO)	PRD 87 (2013) 112001, PRL 113 (2014) 212001
WZ	13	36.1	$\sigma = 51 \pm 0.8 \pm 2.3$ pb	$\sigma = 49.1 + 1.1 - 1$ pb (MATRIX (NNLO))	EPJC 79 (2019) 535
WZ	8	20.3	$\sigma = 24.3 \pm 0.6 \pm 0.9$ pb	$\sigma = 23.92 \pm 0.4$ pb (MATRIX (NNLO))	PRD 93, 092004 (2016)
WZ	7	4.6	$\sigma = 19 + 1.4 - 1.3 \pm 1$ pb	$\sigma = 19.34 + 0.3 - 0.4$ pb (MATRIX (NNLO))	EPJC 72 (2012) 2173
ZZ	13.6	29.0	$\sigma = 16.9 \pm 0.7 \pm 0.7$ pb	$\sigma = 16.7 \pm 0.4$ pb (Matrix (NNLO) & Sherpa (NLO))	PLB 855 (2024) 138764
ZZ	13	36.1	$\sigma = 17.3 \pm 0.6 \pm 0.8$ pb	$\sigma = 16.9 + 0.6 - 0.5$ pb (Matrix (NNLO) & Sherpa (NLO))	PRD 97 (2018) 032005
ZZ	8	20.3	$\sigma = 7.3 \pm 0.4 + 0.4 - 0.3$ pb	$\sigma = 8.284 + 0.249 - 0.191$ pb (NNLO)	JHEP 01 (2017) 099
ZZ	7	4.6	$\sigma = 6.7 \pm 0.7 + 0.5 - 0.4$ pb	$\sigma = 6.735 + 0.195 - 0.155$ pb (NNLO)	JHEP 03 (2013) 128 , PLB 735 (2014) 311
t_{s-chan}	13	140	$\sigma = 8.2 \pm 0.6 + 3.4 - 2.8$ pb	$\sigma = 10.32 + 0.4 - 0.36$ pb (NLO+NNL)	JHEP 06 (2023) 191
t_{s-chan}	8	20.3	$\sigma = 4.8 \pm 0.8 + 1.6 - 1.3$ pb	$\sigma = 5.61 \pm 0.22$ pb (NLO+NNL)	PLB 756 (2016) 228-246
$t\bar{t}W$	13	140	$\sigma = 880 \pm 50 \pm 70$ fb	$\sigma = 745 \pm 52$ fb (NNLO QCD + NLO EW)	JHEP 05 (2024) 131
$t\bar{t}W$	8	20.3	$\sigma = 369 + 86 - 79 \pm 44$ fb	$\sigma = 232 \pm 32$ fb (MCFM)	JHEP 11 (2015) 172
$t\bar{t}Z$	13	140	$\sigma = 860 \pm 40 \pm 40$ fb	$\sigma = 860 + 80 - 90$ fb (NLO + NNLL)	arXiv:2312.04450
$t\bar{t}Z$	8	20.3	$\sigma = 176 + 52 - 48 \pm 24$ fb	$\sigma = 215 \pm 30$ fb (HELAC-NLO)	JHEP 11, 172 (2015)
WWW	13	139	$\sigma = 0.82 \pm 0.01 \pm 0.08$ pb	$\sigma = 0.511 \pm 0.018$ pb (NLO QCD)	PRL 129 (2022) 061803
WWZ	13	79.8	$\sigma = 0.55 \pm 0.14 + 0.15 - 0.13$ pb	$\sigma = 0.358 \pm 0.036$ pb (Sherpa 2.2.2)	PLB 798 (2019) 134913
$t\bar{t}t\bar{t}$	13	140	$\sigma = 22.5 + 4.7 - 3.4 + 6.6 - 5.5$ fb	$\sigma = 13.4 + 1 - 1.8$ fb (NLO QCD + EW)	EPJC 83 (2023) 496

Figure 1.1: Summary of several Standard Model cross-section measurements (a) with associated references (b). The measurements are corrected for branching fractions, compared to the corresponding theoretical expectations.

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. The studies highlighted in this work investigate the performance of RNTuple on the analysis front for ATLAS workflows and serve as documentation for future RNTuple usage.

In the next chapter, I will describe the ATLAS experiment and its detector technology. In Chapter 3, I will introduce the ATLAS software and computing system, and explain our data contents. In Chapter 4, I will give an introduction to RNTuple and TTree. Examples of how RNTuple is applied in comparison to TTree will be shown. In Chapter 5, I will demonstrate the performance studies conducted for RNTuple and how they compare with TTree. In Chapter 6, I will describe the Analysis Grand Challenge (AGC) which served as a benchmark for RNTuple. Finally I will give my conclusions in Chapter 7.

CHAPTER 2

THE ATLAS EXPERIMENT

A Toroidal LHC Apparatus, mainly known as ATLAS, is a particle physics experiment located at the Large Hadron Collider near Geneva, Switzerland. ATLAS is 44 meters long and 25 meters in diameter, making it the largest detector ever constructed for a particle collider. It is designed as a general-purpose particle physics experiment, focused on precision measurements of the Standard Model. Research from ATLAS physicists has led to ground-breaking discoveries, such that of the Higgs boson.

2.1 Detector Technology

To do this, ATLAS has six different detecting subsystems wrapped concentrically in layers around the collision point to record the trajectory, momentum, and energy of particles. Apart, a hurge magnet system bends the paths of the charged particles so that their momenta can be measured as precisely as possible. Overall, the detector tracks and identifies particles to investigate a wide range of physics.

2.1.1 Inner Detector

what is its main functions

2.1.1.1 Pixel Detector

2.1.1.2 Semiconductor Tracker

2.1.1.3 Transition Radiation Tracker

2.1.2 Calorimeter

2.1.2.1 Liquid Argon Calorimeter

2.1.3 Tile Hadronic Calorimeter

2.1.4 Muon Spectrometer

2.1.4.1 Thin Gap Chambers

2.1.4.2 Resistive Plate Chambers

2.1.4.3 Monitored Drift Tubes

2.1.4.4 Small-Strip Thin-Gap

2.1.4.5 Micromegas

2.1.5 Magnet System