

https://dspace.org/

Not a jet all the way: an exploration of the strongly interacting dark sector in ATLAS and beyond

Sinha, Sukanya S

2024
https://hdl.handle.net/10539/37463

Downloaded from a DSpace repository. https://dspace.org/

Department of Physics,
University of Witwatersrand

PhD THESIS

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

Not a jet all the way - an exploration of the strongly interacting dark sector in ATLAS and beyond

Student:
Sukanya Sinha
Student number: 2296508

Supervisor:
Prof. Deepak Kar
Department of Physics

March 17, 2023



DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Sukanya Sinha
17th day of March 2023 at Johannesburg

Abstract

Collider searches for dark matter (DM) so far have mostly focussed on scenarios, where DM particles are produced in association with heavy standard model (SM) particles or jets. However, no deviations from SM predictions have been observed so far. Several recent phenomenology papers have proposed models that explore the possibility of accessing the strongly coupled dark sector, giving rise to unusual and unexplored collider topologies. One such signature is termed as semi-visible jet (SVJ), where parton evolution includes dark sector emissions, resulting in jets interspersed with stable invisible particles. Owing to the unusual MET-along-the-jet event topology this is still a largely unexplored domain within LHC. This thesis presents the first results from a search for SVJ in t-channel production mode in pp collisions for an integrated luminosity of 139 fb^{-1} at centre-of-mass energy corresponding to 13 TeV at the LHC, based on data collected by the ATLAS detector during 2015-2018. Additionally, studies are performed to explore the use of jet substructure methods to distinguish SVJ from SM jets in the first two scenarios, using observables in a IRC-safe linear basis, and ways forward are proposed for this approach to dark-matter at the LHC, including prospects for estimating modelling uncertainties.

Dedicated to my mother, for believing in me and dreaming along with me...

Acknowledgements

It is usually believed that the PhD journey is a solo trip, but I donot think I would be able to embark on, let alone finish this journey without the constant support of the innumerous wonderful people in my life.

To paraphrase Galileo – "you cannot teach a person anything; you can only help them discover it in themselves". My PhD advisor Deepak, did more than just help me discover myself during these years. He went above and beyond to train me, not only to become a rigorous and independent physicist, but to develop my skillsets as a science communicator. He was a strict taskmaster when I was lazy, and a supporting friend during the tough days when nothing seemed to be working. Staying up till late night because his student is still making plots for next day's meeting or going out of his way to make sure my work is recognised, I always knew I could count on him to guide me through my flaws and mistakes. I could keep on writing pages to no end, because words cannot sum up how lucky and grateful I feel to have Deepak as my PhD advisor. I started this journey with him as a student, but finishing it as a collaborator and friend.

I have been lucky enough during my PhD to get funding from the University of Witwatersand Sell schop Grant and subsequently South Africa's National Research Foundation in the form of Extension support scholarship for my doctoral studies. I would like to thank both the organisations for their support.

The help and support I got from Louie, Chris and Zach during my ATLAS authorship qualification task deserves special mention. They helped me navigate through Athena and never lost patience when I had naive issues. I never identified myself as a "coder" but they believed I could get it done; and that trust is truly appreciated.

Doing an ATLAS analysis completely alone can be challenging, but I feel the word "alone" is a misnomer since I got valuable advice and support from several people at each step of my analysis. Sam helped me identify initial challenges when we were starting out the analysis. The analysis editorial board members – Mark, Christian and Marco have been nothing less than supportive and came up with reasonable solutions for the interim problems. Giuliano's help in performing the reinterpretation of the semi-visible jet signals in terms of the current monojet bounds and identifying potential problems as the analysis progressed, was extremely valuable. Finally, Marie-Helene and Peter were excellent readers when the conference note was in track for ICHEP 2022, and gave very useful and detailed feedback that improved the quality of the draft

When it comes to learning, I owe all my knowledge of statistics and associated frameworks to Xifeng. He never got tired of my relentless questions and debugging requests. I could always count on him to come up with the right solution to any statistics problem I encountered and give a very factual answer to even a dumb question; and I cannot thank him enough for everything that he taught me.

Every analysis needs a sharp and intelligent eye that can catch potential flaws and improve the physics result. Aurelio was the sharp and helpful eye my analysis needed. His excellent intuition and constant push to understand why something was not behaving the way it should, has helped me develop a keener eye as well, and I am grateful I got to know such a rigorous physicist from close quarters throughout my PhD time.

Although I was a "COVID era" PhD student, I was extremely fortunate to receive a MCnet Studentship position and visit University of Glasgow in 2021, to work with Andy. Those six months were a life changing experience for me. Andy was a fabulous supervisor, pushing me to test my limits as a coder and I can safely say, I owe all my python skills to him. His enlightening monologues about Physics and life taught me more than he probably would have guessed as a first pass. Although always putting out "fires" on the work front, his constant curiousity to learn something new is awe-inspiring for me.

I would also like to thank the ATLAS Analysis Top reconstruction group and the ATLAS jets and missing energy group for always being there to try and resolve my queries throughout the past three years. They provided very useful suggestions whenever I was stuck technically or conceptually.

Tim deserves special thanks for always being available to answer my naive theoretical questions, and providing a helping hand whenever we were conceptually struggling with the analysis. I knew no matter what dumb question I have, Tim would have a very friendly approach and bring clarity to my thoughts. I would also like to thank Benjamin for everytime I bothered him with general theoretical queries and always received very precise answers that definitely broadened my knowledge.

It is true that you can learn a lot from your peers and I have been very lucky to find friends amongst my peers. I would like to thank Volker and Ian for always lending an ear when I wanted to rant about anything. Those conversations helped in more ways than they can imagine. The ATLAS Early Career Scientist Board kept me sane throughout the two years, when I was serving the board as a member. The people I met during that time continue to be an integral part of my social circle. Flavia has been an amazing colleague and friend during and after our ECSB stint together and I am happy I got to know her.

I would like to thank Caterina for giving me the first opportunity to attend a research workshop in Lund University in 2019, where I was able to interact and exchange ideas with brilliant physicists. The workshop was a significant morale boost for a beginner PhD student like myself. Following that, over the years I have learnt quite a lot from her, just by being present in similar meetings and observing how to maturely tackle any problem that maybe thrown in her direction. I hope to continue learning from her in the near future as well.

Once the physics result is out, it is necessary to get the right platform to share the results. I would like to thank Annapaola for giving me the first opportunity to share my ATLAS result with the wider community at the Semi-visible jets workshop in Zurich. I would also like to thank the organisers of ICHEP 2022, ISMD 2022, BOOST 2022 and EDSU2022 for providing attendance support and allowing me to share my results with the relevant physics communities. I would also like to thank Andrzej for organising a wonderful MCnet school in Zakopane. The interactions I had with people and the lectures I attended, helped me gain a better perspective of topics I don't deal with on a regular basis. SA-CERN has always been accomodating for my numerous trips to CERN and associated ATLAS meetings over the past three years and I am grateful for the opportunity to travel so frequently throughout my PhD, when the world wide situation permitted it.

My research group at Wits — Lawrence, Marvin, Hannah, Danielle and Mvelo — was amazing, and I cherish all the memories I have with them, whether it was a movie night or someone's farewell dinner. My TA group for PHYS1025 was always helpful and I would like to thank Saga and Hannah especially for their willingness to jump in when I had other pressing engagements.

I was fortunate enough to have people around me who inspired and motivated me since my childhood days. I would like to thank my middle school Chemistry teacher, Sanjoy Ghosh for always pushing me towards challenging problems. My high school Physics teacher, Sharmistha Maitra inspired me in many more ways than she can imagine. I am forever grateful to Bis warup Mukhopadhyaya for identifying and recognising the curiousity bug in me, when I was in my final year of Bachelors, and believing me enough to provide a studentship at Harish Chandra Research Institute. This student ship was my first true exposure to academia beyond course work. He maintained his support throughout my master's degree, eventually leading to my master's dissertation. I would like to thank Bhabani Prasad Mandal for being an amazing master's dissertation supervisor and giving me the freedom to tackle a problem of my choice. Sujan was the reason I came to know about a PhD position at Wits and Anirban Kundu played a crucial role in convincing me that coming to Wits for a PhD is a step in the right direction, and I will always be thankful to them for their role in the grand scheme of things.

On my various trips to CERN and other ATLAS / non-ATLAS meetings, I have been immensely fortunate to meet several wonderful people, some of who ended up becoming my close friends. I would like to thank each and everyone of them for being there for me during various stages over the last few years. I would like to thank my mother, my brother and all my relatives for never losing faith in me, and supporting me through thick and thin. My mother and brother would listen to my constant blabbering about how amazing Physics is. I have made them sit through so many of my public talks that I have lost count, but they would always cheer me on. I want to mention Debanjan for being an amazing and caring friend for the past 10 years with many more to come, Akanksha for being my partner in crime and travel since our first trip in Glasgow. Finally, I would like to thank Suvam but this does not need any explanation. He knows what he means to me.

Contents

1 Standard Model - background and drawbacks **17**
- 1.1 A "mandatory" overview of the Standard Model . . . 17
- 1.2 Standard Quantum Chromodynamics – a precursor . . . 23
  - 1.2.1 Strong coupling and asymptotic freedom . . . 24
  - 1.2.2 A non-arduous tour of the fundamentals of QCD inspired event generation . . . 25
- 1.3 Monte Carlo event generators . . . 28
- 1.4 The limitations of SM . . . 32

2 Strongly interacting dark sector - why and how? **33**
- 2.1 Moving on to... the overview of dark matter . . . 33
  - 2.1.1 Early explorations . . . 33
  - 2.1.2 Dark matter relic density . . . 34
  - 2.1.3 Dark matter particle candidates . . . 35
  - 2.1.4 Interactions between Standard Model and dark matter . . . 35
- 2.2 Simplified Models of dark matter . . . 37
- 2.3 QCD-like dark sector . . . 38
  - 2.3.1 Semi-visible jets theory model . . . 39
  - 2.3.2 Hidden Valley Shower . . . 40
  - 2.3.3 Semi-visible jets topology and extra jets . . . 43

3 Collider basics and getting to know the ATLAS detector **45**
- 3.1 Large Hadron Collider at CERN . . . 45
- 3.2 Give me the coordinates . . . 48
- 3.3 Collisions – why do we love them? . . . 49
- 3.4 Hello ATLAS detector . . . 50
  - 3.4.1 Inner detector . . . 51
  - 3.4.2 Electromagnetic Calorimeter . . . 53
  - 3.4.3 Hadronic Calorimeter . . . 54
  - 3.4.4 Muon spectrometer . . . 56
- 3.5 Object reconstruction . . . 57
  - 3.5.1 Jets – how do we make them? . . . 57
  - 3.5.2 $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction . . . 62
- 3.6 Triggering on the objects . . . 63
- 3.7 Detector simulation . . . 66
- 3.8 Existing searches for Dark Matter with jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ in ATLAS . . . 67

4 Search for non-resonant production of semi-visible jets in ATLAS 69
4.1 Context . . . 69
4.2 Standard Model background processes . . . 70
4.3 Object definitions . . . 71
4.4 Dataset and Simulation . . . 73
4.4.1 Simulated samples . . . 73
4.4.2 Data samples . . . 74
4.5 Event selection and cleaning . . . 75
4.5.1 Triggering strategy . . . 75
4.5.2 Data quality . . . 76
4.5.3 Jet Cleaning . . . 76
4.5.4 Event selection . . . 76
4.6 Analysis strategy . . . 79
4.6.1 Inclusive distributions and definition of signal and control regions . . . 80
4.6.2 Signal region distributions . . . 86
4.6.3 Control region distributions . . . 89
4.7 Systematic uncertainties . . . 90
4.7.1 Theoretical systematic uncertainties . . . 90
4.7.2 Experimental systematic uncertainties . . . 92
4.8 Fit strategy and background estimation . . . 93
4.8.1 General strategy and fitting procedure . . . 93
4.8.2 Verification of multijet background estimation . . . 94
4.9 Simulaneous fit of signal region and control region . . . 97
4.9.1 Post-fit distributions in 1LCR . . . 97
4.9.2 Post-fit distributions in 1L1BCR . . . 97
4.9.3 Post-fit distributions in 2L CR . . . 97
4.9.4 Post-fit distribution for combined fit in control region and signal region 98
4.10 Results . . . 103
4.11 Conclusions . . . 105

5 Phenomenological explorations of semi-visible jets 106
5.1 Event generation . . . 106
5.2 Exploring jet substructure of semi-visible jets . . . 107
5.2.1 Analysis strategy . . . 107
5.2.2 Jet substructure observables . . . 108
5.2.3 Results . . . 110
5.2.4 Understanding the model dependence . . . 111
5.2.5 Origin of the differences . . . 112
5.3 Exploring new observables for dark sector . . . 116
5.3.1 Energy flow polynomials . . . 116
5.3.2 Results . . . 117
5.4 Semi-visible jet production with heavy flavour . . . 121
5.4.1 Signal modelling . . . 121
5.4.2 Signal reconstruction . . . 121
5.4.3 Search strategy . . . 123
5.5 Conclusions . . . 126

6 Monte Carlo Truth Classifier **127**
6.1 Necessity of proper truth definitions . . . 127
6.2 Original Monte Carlo Truth Classifier and drawbacks . . . 128
6.3 New implementation of MCTC . . . 129

Bibliography **132**

A Appendices **148**
A.1 Use of reclustered jets in ATLAS SVJ analysis . . . 148
A.1.1 Particle level studies . . . 148
A.1.2 Detector level studies . . . 151
A.1.3 Performance . . . 152
A.2 Tests for non-collisional background . . . 156

List of Figures

1.1 SM particle content . . . 18
1.2 Shape of the potential of a complex scalar field . . . 21
1.3 Running of α_s at LO . . . 25
1.4 Schematic display of IRC safety . . . 26
1.5 Schematic diagram of FSR and ISR shower development . . . 28
1.6 Schematic diagrams of Lund string and cluster hadronisation steps . . . 29
1.7 Potential between a $q\bar{q}$ pair, as a function of distance between them. . . . 30
1.8 Simplified schematic diagram of a complete event generation, with a Z+jets process . . . 31

2.1 Types of DM searches governing the WIMP paradigm . . . 36
2.2 s-channel and t-channel production modes for DM production. . . . 39
2.3 Diagram showing the direction of $E_{\mathrm{T}}^{\mathrm{miss}}$ for different r_{inv} values . . . 40
2.4 Schematic diagram showing evolution of jets with respect to different r_{inv} scenarios. . . . 43
2.5 An example of a Feynman diagram showing cross contributions during t-channel production . . . 44

3.1 A schematic diagram of the CERN accelerator complex . . . 46
3.2 Luminosity deliverance distributions . . . 47
3.3 Number of Interactions per Crossing . . . 48
3.4 Diagram showing the coordinate system used in the LHC . . . 49
3.5 Schematic illustration of different types of inelastic collisions . . . 50
3.6 Schematic diagram of a generic detector and ATLAS . . . 51
3.7 Schematic representation of the ATLAS inner detector . . . 52
3.8 Schematic representation of the ATLAS calorimeter . . . 55
3.9 Schematic representation of the ATLAS muon spectrometer . . . 56
3.10 Illustration of calorimeter-only against particle-flow configurations . . . 58
3.11 Topoclusters . . . 58
3.12 Particle-flow algorithm . . . 59
3.13 anti-k_t, k_{T} and Cambridge-Aachen jets . . . 61
3.14 Schematic illustration of successive triggering steps . . . 64
3.15 An example trigger turn-on curve plot . . . 64
3.16 The flow of the ATLAS simulation software . . . 66
3.17 Schematic diagram showing how $E_{\mathrm{T}}^{\mathrm{miss}}$ is determined in the transverse cross-section of a LHC detector . . . 67
3.18 Monojet reinterpretation of semi-visible jet signals . . . 68

4.1 Illustrative Feynman diagram and subsequent production mechanism of semi-visible jets via t-channel . . . 69

4.2 Inclusive kinematic distributions for four benchmark signals with different r_{inv} fractions . . . 81
4.3 Inclusive kinematic distributions for four benchmark signals with different mediator masses . . . 82
4.4 Inclusive kinematic distributions for four benchmark signals and background with different r_{inv} fractions . . . 83
4.5 Inclusive kinematic distributions for four benchmark signals and background with different mediator masses . . . 84
4.6 Correlation between the H_{T} and $E_{\text{T}}^{\text{miss}}$ distributions . . . 85
4.7 Correlation between the p_{T} balance and maxminphi distributions . . . 87
4.8 Comparisons of shape of $p_{\text{T}}^{\text{bal}}$ and $|\phi_{\text{max}} - \phi_{\text{min}}|$ distributions . . . 88
4.9 The definition of the 9-bins in $|\phi_{\text{max}} - \phi_{\text{min}}|$ and $p_{\text{T}}^{\text{bal}}$, defined identically in SR, VR and in each CR. . . . 89
4.10 Comparisons of shape of $p_{\text{T}}^{\text{bal}}$ and $|\phi_{\text{max}} - \phi_{\text{min}}|$ distributions between the total background and data for 1LCR. . . . 89
4.11 Comparisons of shape of $p_{\text{T}}^{\text{bal}}$ and $|\phi_{\text{max}} - \phi_{\text{min}}|$ distributions between the total background and data for 1L1BCR. . . . 90
4.12 Comparisons of shape of $p_{\text{T}}^{\text{bal}}$ and $|\phi_{\text{max}} - \phi_{\text{min}}|$ distributions between the total background and data for 2LCR. . . . 90
4.13 Comparison of different samples of $t\bar{t}$ for ME and PS variations . . . 91
4.14 Comparison for DR and DS scheme for tW process modelling . . . 92
4.15 Kinematic distributions for multijet background estimation . . . 96
4.16 9 bin fitted histogram in $E_{\text{T}}^{\text{miss}}$ vs H_{T} in 1LCR . . . 97
4.17 9 bin fitted histogram in $E_{\text{T}}^{\text{miss}}$ vs H_{T} in 1L1BCR . . . 98
4.18 9 bin fitted histogram in $E_{\text{T}}^{\text{miss}}$ vs H_{T} in 2LCR . . . 98
4.19 9 bin fitted histogram in $E_{\text{T}}^{\text{miss}}$ vs H_{T} in SR . . . 99
4.20 Dominant systematic uncertainty ranking in CR-SR combined fit . . . 101
4.21 Correlation matrix for CR-SR combined fit . . . 102
4.22 Postfit distributions of unblinded SR kinematic variables . . . 103
4.23 Exclusion limit plots for different signals with r_{inv} fractions. . . . 104
4.24 Upper limit on coupling strength . . . 104

5.1 $\eta - \phi$ distributions for large-radius jets . . . 107
5.2 Distributions of the azimuthal angle difference between the leading and sub-leading jets . . . 108
5.3 Distributions of $E_{\text{T}}^{\text{miss}}$ and leading jet p_{T} for different signals . . . 108
5.4 Comparison of substructure observables . . . 110
5.5 Comparison of substructure observables for checking model dependence . . . 112
5.6 Comparison of substructure observables by clustering dark hadrons in final state . . . 113
5.7 Comparison of substructure observables to interpret behaviour of dark hadrons 114
5.8 Comparison of substructure observables with intermediate and final state dark hadrons . . . 114
5.9 Comparison of substructure observables to check the effect of trimming . . . 115
5.10 Comparison of substructure observables clustering only the visible hadrons and clustering also with final dark hadrons. . . . 115
5.11 EFP construction: vertex and angular connectors . . . 116
5.12 EFP construction: a degree one polynomial . . . 116
5.13 EFP diagram with 4 constituents and 5 angularity connectors . . . 117

5.14 LLR summary distribution containing 10 EFP diagrams . . . 118
5.15 EFP distributions corresponding to spikes in LLR summary plot. . . . 118
5.16 Comparison of known jet substructure observables with selected EFPs . . . 119
5.17 Comparison of other known jet substructure observables with selected EFPs 120
5.18 b-tagged jet multiplicity in signal . . . 123
5.19 $\eta - \phi$ distributions for variable-radius jets . . . 123
5.20 The correlation of charged lepton p_{T} against the $\Delta\phi$(closest jet, $E_{\mathrm{T}}^{\mathrm{miss}}$)distance from the closest b-tagged jet . . . 124
5.21 Kinematic distributions for signal and leading background processes . . . 125
5.22 Kinematic distributions for signal and leading background processes after SR selections . . . 125

6.1 Flowchart of the old MCTC classification scheme. . . . 129
6.2 Flowchart of the new MCTC classification scheme. . . . 130

A.1 Particle level distribution of area normalised H_{T}, for R15 reclustered jets for varying M_ϕ and r_{inv}. . . . 148
A.2 Particle level objects plotted in the $\eta - \phi$ plane . . . 149
A.3 Particle level distributions for different mediator mass and r_{inv} fractions . . . 150
A.4 Particle level distributions of average overlap between reclustered jets of different radius parameter . . . 150
A.5 Reconstruction level objects plotted in the $\eta - \phi$ plane for three events . . . 151
A.6 Reconstruction level objects plotted in the $\eta - \phi$ plane for three other events. 152
A.7 $p_{\mathrm{T}}^{mean}, p_{\mathrm{T}}^{rms}, \eta^{mean}$ and η^{rms} distributions for truth-matched reco-level jets for signal point r_{inv}= 0.2. . . . 153
A.8 $p_{\mathrm{T}}^{mean}, p_{\mathrm{T}}^{rms}, \eta^{mean}$ and η^{rms} distributions for truth-matched reco-level jets for signal point r_{inv}= 0.4. . . . 153
A.9 $p_{\mathrm{T}}^{mean}, p_{\mathrm{T}}^{rms}, \eta^{mean}$ and η^{rms} distributions for truth-matched reco-level jets for signal point r_{inv}= 0.6. . . . 154
A.10 $p_{\mathrm{T}}^{mean}, p_{\mathrm{T}}^{rms}, \eta^{mean}$ and η^{rms} distributions for truth-matched reco-level jets for signal point r_{inv}= 0.8. . . . 154
A.11 Number of subjets seeding the reclustered jet (left). This leads to weird trend in JMS for low p_{T} and mass regime (right). . . . 155
A.12 Effect of NCB cleaning on data and MC. . . . 156

List of Tables

2.1 List of Hidden Valley particles . . . 41
2.2 PYTHIA8 HV parameter choices. . . . 43

4.1 Summary of jet reconstruction criteria. . . . 71
4.2 Summary of b-tagging selection criteria. . . . 72
4.3 Summary of $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction criteria. . . . 72
4.4 Overview of the overlap removal between objects and the corresponding matching criteria, listed according to priority. . . . 73
4.5 Details of signal samples generated with r_{inv} values of 0.2, 0.4, 0.6 and 0.8, with dark hadron mass of 10 GeV, each with full detector simulation. . . . 74
4.6 Summary of generators used for simulation of background processes, along with the PDF and tune used. . . . 74
4.7 Cutflow table for four benchmark signals of different mediator masses in and r_{inv} of 0.4 . . . 77
4.8 Cutflow table for four benchmark signals of mediator mass 2000 GeV with different r_{inv} fractions . . . 77
4.9 Summary of event pre-selections for different SR and CRs. . . . 79
4.10 Signal contamination values for four benchmark signals of different mediator masses and r_{inv} of 0.4, for determining H_{T} threshold for CR . . . 85
4.11 Signal significance values for four benchmark signals of different mediator masses and r_{inv} of 0.4, for determining H_{T} threshold for SR . . . 86
4.12 Signal contamination values for four benchmark signals of different r_{inv} fractions, for determining H_{T} threshold for CR . . . 86
4.13 Signal significance values for four benchmark signals of different r_{inv} fractions, for determining H_{T} threshold for SR . . . 86
4.14 Signal significance values for signals of mediator mass 3.5 TeV and r_{inv} of 0.6, for the $p_{\mathrm{T}}^{\mathrm{bal}}$ bin optimisation for deciding final 9 bin grid. . . . 87
4.15 Signal significance values for signals of mediator mass 3.5 TeVand r_{inv}0.6, for the maxminphi bin optimisation for deciding final 9 bin grid. . . . 88
4.16 Signal significance values for signals of mediator mass 3.5 TeVand r_{inv} 0.6, for for the $p_{\mathrm{T}}^{\mathrm{bal}}$ and maxminphi bin grid used for fitting. . . . 88
4.17 Post-fit yields from background-only fit . . . 99
4.18 Scale factors for each background processes . . . 100

5.1 Hidden Valley model parameters considered in the study . . . 111
5.2 Grid formation, translation Figure 5.13 into a set of "particle" pairs. . . . 117
5.3 Cutflow table summarising selections used in the study . . . 125

Preface

Standard Model (SM) of particle physics has seen several successes throughout the past century as discussed in Chapter 1. However, the existence of dark matter as validated by astrophysical studies prompted an increase in the number of search programmes exploring the Beyond Standard Model sector using Large Hadron Collider (LHC) data. The obvious question to ask - Is dark matter being produced at the very high energy collisions produced at the LHC? If so, what is its signature? The only SM particle which does not interact with any of the detector components is the neutrino, since they are colour- and charge-neutral, and have faint weak interactions. Hence, the presence of neutrinos causes an imbalance of transverse momentum, which is termed as missing transverse momentum. However, if a collision only produces DM particles, we would not see anything in the detector. So we probe what is termed as mono-X signatures, where X is any SM particle (or object, like a jet) being produced along with a DM particle. The unbalanced production of X is expected to yield a large missing transverse momentum signature, inconsistent with neutrinos. To date, dark matter searches at the LHC have usually focused on Weakly Interacting Massive Particles (WIMPs), but since the existing searches in colliders have found no evidence of DM so far, several recent phenomenology papers have explored the possibility of accessing the dark sector with unique collider topologies.

My primary doctoral research spanned across experimental and phenomenological studying a class of dark sector models known as dark Quantum Chromodynamics (dark QCD) models. In these models, dark matter is a composite particle emerging from a new "dark force" that is similar to the strong force in the SM, governed by the theory of Quantum Chromodynamics (QCD). As the strong force is responsible for the interactions between the constituents of the proton (quarks and gluons), the new "dark force" would explain interactions between the constituents of dark matter particles, and between these constituents and known matter. The details of this model have been discussed in Chapter 2.

Running a multi-purpose detector like ATLAS involves building detector components and monitoring their operation, translating detector signals into physics information, obtaining calibrations, and managing the computing resources, amongst other tasks, and one individual cannot contribute to every component of such a large scale experiment. Chapter 3 discusses the basics of collider physics and the details of the ATLAS detector system.

Semi-visible jets (SVJ) are jet-like collider objects where the visible states are interspersed with DM particles. Recent phenomenological studies have targetted a non-WIMP scenario which eventually leads to interesting collider signatures, where the final state consists of a mixture of stable, invisible dark hadrons and visible hadrons from the unstable subset of dark hadrons that promptly decay back to SM particles. The total momentum of the dark matter is hence correlated with the momentum of the visible states, leading to the event missing transverse energy close to a jet. As this is also a signature of jet mismeasurement in a detector due to presence of dead regions, or measurement of fake $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions, especially from multijet processes, this class of jets had so far been unexplored in ATLAS, and such events had typically been ignored. However, my main doctoral analysis – exploring SVJ

in t-channel production mode, discussed in Chapter 4 – has set the first ever limits on this yet unexplored phase space. This has opened up several avenues of accessing the strongly interacting dark sector, by analysing the phenomenological consequences of the observed results.

The SVJ signature was proposed a few years back, however the theoretical papers did not include a detailed collider search strategy. I worked on how best we can identify these unconventional jet signatures from the ones arising from Standard model processes and all the studies have been discussed in Chapter 5. One way of distinguishing SVJ from standard jets, is by looking at the internal structure of the jet, and studying the differences in jet behaviour, when dark matter is assumed to be contributing to their formation. I proposed ways to look for such unique collider signatures, using jet-substructure (JSS) observables. I have studied several jet substructure observables to compare SVJ and light quark or gluon initiated jets. Looking at the different dark hadron fractions allows us to check whether the substructure is created by the interspersing of visible hadrons with dark hadrons, or from certain model dependencies. Similarly, since this class of jets are unconventional by nature, there can be possibilities where new JSS observables might be more sensitive to these jets compared to light quark/gluon jets. As a MCnet early-career researcher, I have explored the option of designing new observables in a IRC-safe linear basis in the RIVET framework, using energy flow polynomials and this study proposed ways forward for this approach to dark-matter at the LHC, including prospects for estimating modeling uncertainties.

Searching for new physics is important, however understanding the flaws in present physics process modeling and trying to improve our current understanding is equally important. Hence, apart from the analysis and related performance component, I redeveloped the Monte Carlo Truth Classifier (MCTC) tool as part of my ATLAS Qualification task as discussed in Chapter 6, which involved improving the existing truth level particle definitions used within the collaboration by eliminating generator dependence in the classification of truth particles.

List of papers/pre-prints and public results

- Towards discrimination and improved modelling of dark sector showers, S. Sinha et al, arXiv:2209.14964 [hep-ph]
- 2B or not 2B, a study of bottom-quark-philic semi-visible jets, S. Sinha et al, arXiv:2207.01885 [hep-ph]
- Search for semi-visible jets in t-channel production mode using Run-2 ATLAS data, ATLAS collaboration, ATLAS-CONF-2022-038
- Theory, phenomenology, and experimental avenues for dark showers: a Snowmass 2021 report, S. Sinha et al, arXiv:2203.09503 [hep-ph]
- Hitting two BSM particles with one lepton-jet: search for a top partner decaying to a dark photon, resulting in a lepton-jet, S. Sinha et al, 10.21468/SciPostPhys.13.2.018
- Exploring Jet Substructure in Semi-visible jets, D. Kar and S. Sinha, 10.21468/SciPostPhys.10.4.084
- Constraining the Dark Sector with the monojet signature in the ATLAS experiment, ATLAS collaboration, ATL-PHYS-PUB-2021-020

Chapter 1

Standard Model - background and drawbacks

The Standard Model (SM) of particle physics [1] has been successful in explaining almost all the experimentally observed sub-atomic phenomena. It intricately explains the existence of 17 fundamental particles in nature, starting from the three generations of fermions (quarks and leptons), to the vector (g , γ, W, Z) and scalar (H) bosons. Despite the fact that SM is highly robust, it is not without its limitations, however it still remains an useful starting point in order to probe for new physics. This chapter intends to summarise the SM and state some of the limitations, which will pave the way for the rest of the thesis.

1.1 A "mandatory" overview of the Standard Model

The SM of particle physics describes the fundamental particles that comprise everything in the visible universe and their interactions, barring the gravitational interactions between matter. The first step towards the formal construction of the SM dates back to S. Glashow's efforts in 1961, to connect two of the fundamental forces of nature – electromagnetic (EM) force and weak nuclear force [2]. The SM explains the existence of massive particles by the Higgs mechanism, where a spontaneously broken symmetry associated with a scalar field (Higgs field) results in the appearance of mass terms for gauge bosons and fermions, originally proposed by P. Higgs, R. Brout, F. Englert, G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble in 1964 [3–5]. This was followed by the contributions of S. Weinberg and A. Salam in 1967, which presents the model in its current form [6, 7].

The different particles of the SM are summarised in Fig. 1.1. According to this model, matter is made up of two types of particles – fermions and bosons. Fermions have half-integer spin and obey the principles of Fermi-Dirac statistics. They are known to contribute to the formation of matter as we know it. Bosons are particles with integer spin, obeying the principles of Bose-Einstein statistics, and act as the mediating forces that control the interactions between fermions.

The theory also demands that every particle has an associated anti-particle, which is identical in every quantum property apart from the charges. There are hypothesized exceptions where a particle might be its own anti-particle and such particles are termed as Majorana fermions.

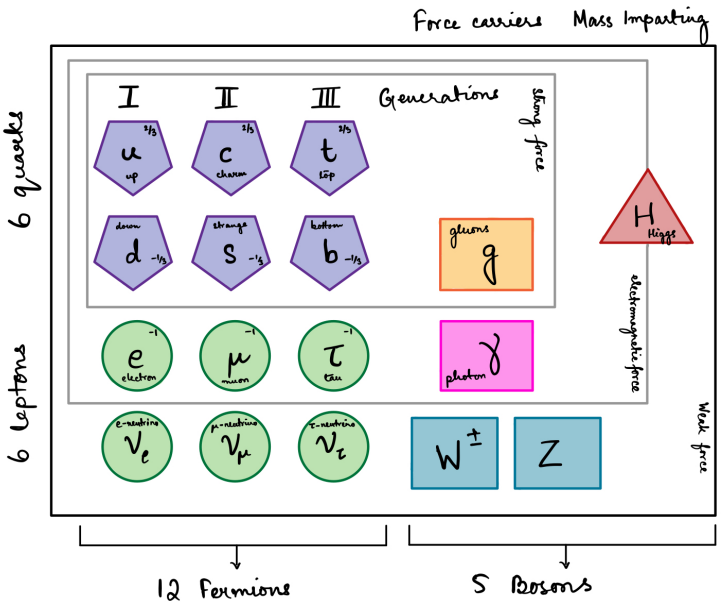


Figure 1.1: An overview of SM. The violet and green coloured blocks represent the three generations of quarks and leptons respectively. The orange block is the gluon, the pink block is the photon and the seagreen blocks are the vector gauge bosons. The red block is the scalar Higgs boson.

Fermions

Fermions are sub-divided into three generations of quarks and leptons, with each generation having a significantly larger mass compared to the previous generation, while the other quantum properties remain unchanged. Every generation consists of a pair of quarks and a pair of leptons, along with their associated anti-particles.

The pair of quarks in each generation are termed as *up-type* and *down-type*, following the naming convention of the first generation quarks. The *up-type* quarks (or anti-quarks) have a charge of $\frac{2}{3}$ (or -$\frac{2}{3}$), whereas the *down-type* quarks (or anti-quarks) have a charge of -$\frac{1}{3}$ (or $\frac{1}{3}$).

According to quantum chromodynamics (QCD), quarks also possess a property called colour charge. There are three types of colour charges – red (r), blue (b), and green (g), with each of them being complemented by three anti-colour charges – anti-red ($\bar{r}$), anti-blue ($\bar{b}$), and anti-green ($\bar{g}$). Every quark (anti-quark) is assumed to carry a colour (anti-colour).

However, due to the presence of colour confinement effect [8], quarks cannot exist as isolated stable particles, and instead combine with other quarks to form colour-neutral states called hadrons. Hadrons can be classified as baryons or mesons, depending on whether they are formed of three quarks or a quark – anti-quark pair. Despite the fact that the probability of particles coupling via strong interactions is orders of magnitude higher than them coupling via weak interactions [1], the range of the strong force is restricted to the approximate size of the proton (i.e. 10^{-15} m), owing to colour confinement constraints.

The pair of leptons in each generation, on the other hand, exist as a charged lepton (anti-lepton) carrying a charge of 1 (-1) and a anti-neutrino (neutrino), which is EM neutral. Leptons also carry weak isospin, and hence interact with themselves and other particles, respectively via the EM and weak forces. The charged lepton masses are again seen to increase with increasing generation, but the case of neutrino masses is a more complicated

[1]The variation of the coupling with the energy scale will be discussed in detail in the next chapter

scenario, since the SM traditionally considers them to be massless. However, it has been inferred from experiments studying neutrino oscillations, that neutrinos have a non-zero, albeit infinitesimally small mass value [9].

Bosons

The SM bosons account for the different fundamental forces, with a broad classification of being either a gauge boson or a scalar boson. The gauge bosons are spin one force carriers, mediating interactions between different particles, whereas scalar bosons are spin zero particles.

- The gluon is a massless gauge boson which mediates the strong force. It couples to quarks (since they contain a colour charge) and also to itself (since it carries a colour charge as well). Gluons mediate as well as participate in strong interactions, thereby allowing the formation of three/four-gluon vertices. There are eight independent colour combinations of gluons possible, as suggested by QCD. One commonly used list of the colour states are:

$$(r\bar{b} + \bar{r}b)/\sqrt{2} \qquad -i(r\bar{b} - \bar{r}b)/\sqrt{2}$$

$$(r\bar{g} + \bar{r}g)/\sqrt{2} \qquad -i(r\bar{g} - \bar{r}g)/\sqrt{2}$$

$$(b\bar{g} + \bar{b}g)/\sqrt{2} \qquad -i(b\bar{g} - \bar{b}g)/\sqrt{2}$$

$$(r\bar{r} - \bar{b}b)/\sqrt{2} \qquad (r\bar{r} + \bar{b}b - 2g\bar{g})/\sqrt{6}$$

- The photon (γ) is another massless gauge boson which mediates the EM force. It couples to all fermions containing a non-zero EM charge (i.e. both quarks and leptons). Unlike gluons, the γ does not contain an EM charge itself.

- The $W^{\pm}$ and Z^0 gauge bosons mediate the weak force. The Z^0 boson is its own anti-particle and can mediate the weak neutral current. Because of its EM neutral nature, it couples to the same fermions as the photon, but also additionally couples to neutrinos. The W^- boson is the anti-particle of the W^+ boson and vice-versa. They have a EM charge of ± 1 and can mediate generation and flavour changing weak processes.

- The "Higgs" boson is the only known scalar boson, which is electrically neutral and is responsible for imparting mass to the other fermions. SM particles acquire mass via their interactions with the Higgs field, and the more massive the particle, the stronger the coupling is with the Higgs boson. The Higgs boson is known to interact with only massive SM particles, and also generates the masses, owing to the fact that $W^{\pm}$ and Z^0 gauge bosons are not massless.

Gauge theories

The SM is a quantum field theory, where the particles can be represented as excitations of quantum fields. These quantum fields of SM and their interactions can be explained using the Lagrangian densities formalism, which are functions of fields and their derivatives respectively. Within the SM, the interactions are described by gauge theories, i.e. the

Lagrangian density is invariant under a set of transformations, which are the symmetries of the theory. These transformations have representation matrices which are the generators of the associated symmetry group.

The SM symmetry groups are $SU(3)_C$ x $SU(2)_L$ x $U(1)_Y$. QCD arises from the $SU(3)_C$ component, containing $(3)^2$ - 1 = 8 generators (representing the eight gluons), whereas quantum electrodynamics (QED) and weak interactions arises from the $SU(2)_L$ x $U(1)_Y$ component of the theory, containing $((2)^2$ - 1) + 1 = 4 generators (representing the γ, $W^{\pm}$ and Z^0 bosons).

Symmetries are important in nature, but symmetry breaking is also important. Explicit symmetry breaking of a theory arises because of an explicit term in the Lagrangian, which breaks the symmetry of the theory actively. In this case the Lagrangian in question contains a symmetry breaking term from which the equations of motion are derived. On the other hand, in the case of spontaneous symmetry breaking (SSB), the vacuum breaks the symmetry. The Lagrangian and corresponding equations of motion still obey the spontaneously broken symmetry.

Necessity of SSB in gauge theories SSB has profound applications in the realm of gauge theories of particle physics. Owing to the short range nature of weak interactions, the theory must necessarily have massive force mediators (the gauge bosons). Adding explicit mass terms for the gauge bosons would violate gauge symmetry. Such an explicit symmetry breaking compromises the unitarity of the theory, and hence, we require some other form of symmetry breaking that not only gives us the masses but also renders the theory renormalizable. To this end the phenomenon of SSB (to be attributed subsequently to Higgs Mechanism) appears as a possible candidate. The Glashow-Weinberg-Salam theory of weak interactions, which relies on the SSB of $SU(2)_L$ x $U(1)_Y$, is renormalizible [10]. In 1971, G. 't Hooft had explicitly showed that gauge theories are renormalizible even in the presence of spontaneous symmetry breaking [11]. This established the importance of spontaneous symmetry breaking in the gauge theory of weak interactions and for that matter in any other gauge theory.

To understand the question of how SM particles are imparted with mass, it is necessary to peek into the Higgs mechanism. The SM Lagrangian consists of the following components:

$$\mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_{fermion} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \tag{1.1}$$

The Weinberg-Salam model consists of three $SU(2)_L$ gauge bosons, W^i_μ where i = 1,2,3, and one $U(1)_Y$ gauge boson, B_μ. The kinetic terms of the theory are,

$$\mathcal{L}_{gauge} = -\frac{1}{4} W^i_{\mu\nu} W^{i\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.2}$$

where,

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk} W^j_\mu W^k_\nu \tag{1.3}$$

is the gluonic field-strength tensor, and

$$B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu \tag{1.4}$$

is the hypercharge field-strength tensor. On introducing a $SU(2)$ doublet of complex scalar fields,

$$\phi = \begin{pmatrix} \phi_1 + i\phi_2 \\ H^0 + i\phi_3 \end{pmatrix} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.5}$$

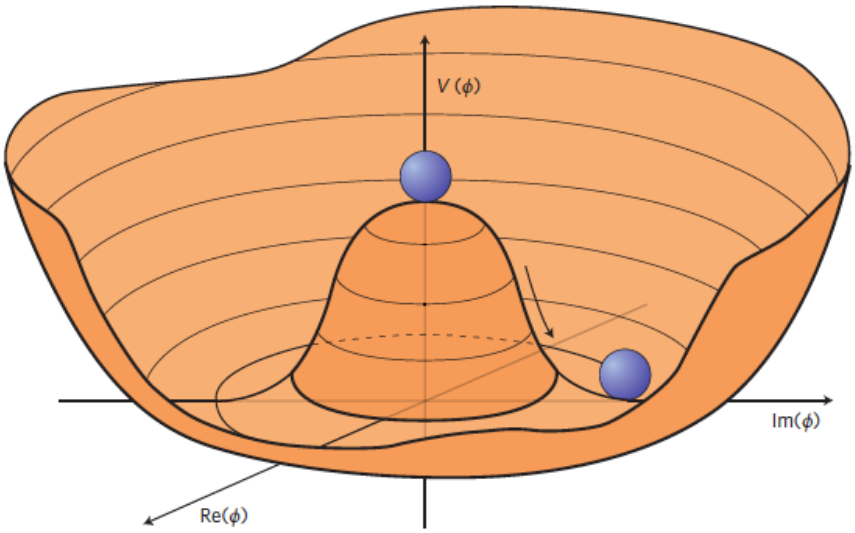


Figure 1.2: Shape of the potential of a complex scalar field, with infinite number of degenerate minima (i.e. vacuum) rotated by the phase rotation, leading to SSB [12].

This is the SU(2) complex Higgs doublet with $SU(2)$ x $U(1)$ invariant scalar potential,

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{1.6}$$

with $\mu^2 < 0$ and $\lambda > 0$. When $\mu^2 < 0$, the scalar field develops a non-zero vacuum expectation value and this leads to SSB. Since there are an infinite number of degenerate minima rotated by the phase rotation, the shape of the potential takes the form of a Mexican hat as shown in Fig 1.2. The λ term gives quartic self-interactions among the scalar fields and for vacuum stability $\lambda > 0$. Now, on minimizing the potential in 1.6 we get,

$$\langle\phi^\dagger\phi\rangle = \frac{-\mu^2}{\lambda} \tag{1.7}$$

where $\langle\phi^\dagger\phi\rangle$ denotes the vacuum expectation value, VEV. The symmetry of the theory enables us to choose

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix} \tag{1.8}$$

via SU(2) $\times$ U(1) transformations. Here, $H^0 = H + v/\sqrt{2}$. With this choice, the scalar doublet has $U(1)_Y$ charge (hypercharge) $Y_\phi = 1$. To conserve electric charge, only a neutral scalar field can acquire a VEV. Thus, ϕ^0 is interpreted to be the neutral component of the doublet. The electromagnetic charge is $Q < \phi >= 0$ and hence EM is unbroken by the scalar VEV to yield the following symmetry breaking scheme: $SU(2)_L$ x $U(1)_Y \rightarrow U(1)_{EM}$

The scalar Higgs part of the Lagrangian is,

$$\mathcal{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) \tag{1.9}$$

To make the $SU(2)_L$ x $U(1)_Y$ symmetry local, gauge fields, W^i_μ for $SU(2)_L$ and B_μ for $U(1)_Y$ have been introduced. The covariant derivative for this weak doublet is,

$$D_\mu = \partial_\mu + ig\frac{\tau_i}{2}W^i_\mu + i\frac{g'}{2}B_\mu Y \tag{1.10}$$

with Y = 1. Here, g and g' are the $SU(2)_L$ and $U(1)_Y$ gauge couplings respectively, and τ_i are the usual Pauli matrices [9].

In unitary gauge, there are no Goldstone bosons [2], thus, only the physical Higgs scalar is present in the spectrum after spontaneous symmetry breaking takes place. Therefore, we can write the scalar doublet in the unitary gauge as follows,

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v+H\end{pmatrix} \tag{1.11}$$

Here, the VEV of the theory, $v = \sqrt{\mu^2/\lambda}$ and H is the real scalar Higgs boson with mass $m_H = \sqrt{2\lambda}v$. It is important to note that the other three degrees of freedom, ϕ_1, ϕ_2, ϕ_3, are absorbed by the weak bosons as shown below.

From the kinetic term of the Higgs potential, we have,

$$(D_\mu\phi)^\dagger(D^\mu\phi) \to \frac{1}{2}(0\ \ v)|\partial_\mu + ig\frac{\tau_i}{2}W_\mu^i + i\frac{g'}{2}B_\mu|^2\begin{pmatrix}0\\ v\end{pmatrix} \tag{1.12}$$

On solving the above,

$$(D_\mu\phi)^\dagger(D^\mu\phi) \to \frac{v^2}{8}\left[g^2\left[(W_\mu^1)^2 + (W_\mu^2)^2\right] + (gW_\mu^3 - g'B_\mu)^2\right] \tag{1.13}$$

Additionally, the square of the covariant derivative involves three and four-point interactions between the gauge and scalar fields. The charged vector boson, W_μ^- and it's complex conjugate is defined as,

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \tag{1.14}$$

Then the g^2 term becomes $\frac{1}{2}(\frac{gv}{2})^2 W_\mu^+ W^{-\mu}$ and yields the W mass,

$$m_W = \frac{gv}{2} \tag{1.15}$$

The two remaining neutral gauge bosons, Z and A, can be defined as,

$$Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}}, \qquad A_\mu = \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2+g'^2}} \tag{1.16}$$

from diagonalizing the mass matrix, thereby giving masses,

$$m_Z = \frac{v}{2}\sqrt{g^2+g'^2}, \qquad m_A = 0 \tag{1.17}$$

This is how the W and Z bosons obtain mass from the Higgs mechanism, whereas the photon remains massless. Counting the degrees of freedom, in the original theory, the complex doublet had four degrees of freedom. After the Higgs mechanism, the scalar Higgs field has 1 degree of freedom, and the three massive weak bosons, $W^\pm$ and Z have three, thereby conserving the number of degrees of freedom.

In order to find out the fermion masses, it is convenient to write the fermions in terms of their left-handed and right-handed projections,

$$\psi_{L,R} = \frac{1}{2}(1 \mp \gamma^5)\psi \tag{1.18}$$

[2]Goldstone bosons appear in theories exhibiting spontaneous breaking of continuous symmetries. Each broken direction of a symmetry gives rise to one such massless scalar.

Here, $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ is the fifth Dirac matrix generally used to project a Dirac field into its left-/right-handed components, with $\gamma^\mu = \gamma^0, \gamma^1, \gamma^2, \gamma^3$ being the four contravariant gamma matrices.

Let us consider the example of electron and its neutrino. Since the W boson couples only with left-handed fermions, thus, the $SU(2)_L$ doublet is constructed as

$$L_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \tag{1.19}$$

The hypercharge in this case is, Y_L = -1. Experimentally, it is known that the W boson does not interact with right-handed fields, thus the right-handed electron must be a $SU(2)_L$ singlet, and has Y_R = -2. Now, considering these hypercharges, the leptons can be coupled to the $SU(2)_L$ x $U(1)_Y$ gauge fields in a gauge invariant manner as follows,

$$\mathcal{L}_{lepton} = i\bar{e_R}\gamma^\mu(\partial_\mu + i\frac{g'}{2}Y_e B_\mu)e_R + i\bar{L_L}\gamma^\mu(\partial_\mu + i\frac{g'}{2}Y_l B_\mu + i\frac{g}{2}\tau_i W^i_\mu)L_L \tag{1.20}$$

Similarly, all the fermions can be included in the SM. A Dirac fermion mass term takes the form,

$$\mathcal{L}_{mass} = -m\bar{\psi}\psi = -m(\bar{\psi_L}\psi_R + \bar{\psi_R}\psi_L) \tag{1.21}$$

The gauge invariant Yukawa coupling of the Higgs boson to the up and down quarks are,

$$\mathcal{L}_d = y_d\bar{Q_L}\Phi d_R + y_u\bar{Q_L}\epsilon\Phi^* u_R + h.c. \tag{1.22}$$

where, $y_{u,d}$ are the 3×3 complex matrices, ϵ is the 2×2 antisymmetric tensor, $\bar{Q_L}$ are the left-handed quark doublets, Φ is the Higgs field, h.c is the Hermitian conjugate, and u_R and d_R are the right-handed up-type and down-type quark singlets respectively, in the weak-eigenstate basis.

This gives the mass matrices of the down-type and up-type quark as,

$$m_d = \frac{Y_d v}{2}, \qquad m_u = \frac{Y_u v}{2} \tag{1.23}$$

Similar couplings can be used to generate the mass terms for other charged leptons, whereas neutrinos remain massless because of the absence of right-handedness.

1.2 Standard Quantum Chromodynamics – a precursor

As discussed briefly in the previous section, QCD is the gauge theory that explains the strong interactions of coloured quarks and gluons. The QCD Lagrangian is given by,

$$\mathcal{L} = \sum_q \bar{\psi_{q,a}}(i\gamma^\mu\partial_\mu\delta_{ab} - g_s\gamma^\mu t^C_{ab}\mathcal{A}^C_\mu - m_q\delta_{ab})\psi_{q,b} - \frac{1}{4}F^A_{\mu\nu}F^{A\mu\nu} \tag{1.24}$$

Here, $\psi_{q,a}$ are the quark-field spinors for quark flavour q, colour index a (a = 1,2,3) since QCD is the SU(3) component of SU(3) × SU(2) × U(1) group, and mass m_q. γ^μ are the Dirac γ-matrices, g_s ($= \sqrt{4\pi\alpha_s}$) is the QCD coupling constant, and $\mathcal{A}^C_\mu$ corresponds to gluon fields with C = 1 - 8 (since, N_C^2 - 1 = 8), i.e. there are 8 gluons. t^C_{ab} are eight 3 × 3 matrices which are the generators in the fundamental representation of Lie algebra of SU(3) group,

and contain information of how a gluon interacting with a quark can rotate the colour of the quark in SU(3) space. In the above equation, the repeated indices are summed over [13–15].

$F^A_{\mu\nu}$ is the field-strength tensor given by,

$$F^A_{\mu\nu} = \partial_\mu \mathcal{A}^A_\nu - \partial_\nu \mathcal{A}^A_\mu - g_s f_{ABC} \mathcal{B}^A_\mu \mathcal{A}^C_\nu \qquad [t^A, t^B] = i f_{ABC} t^C \tag{1.25}$$

with f_{ABC} being the structure constants of SU(3) group.

Looking into QCD from a hadron collider physics perspective, the immediate consequence is that some observables of QCD are calculable using the fundamental parameters of the Lagrangian. However, there are other observables that have to be expressed through models or functions whose effective parameters can only be constrained by fitting to data. Due to the fact that currently there is limited understanding of non-perturbative effects in hadron collisions, this section has been restricted to focus on some key aspects of perturbative QCD (pQCD) that pertain to collider physics [16].

The hard scattering processes in hadron collisions can be described using the collinear factorisation theorem in QCD [17], whereby colliding protons are treated as a collection of partons (point-like constituents of a hadron), carrying a fraction of the proton energy. Parton distribution functions (PDFs) [18] are probabilities of finding these partons with definite energy fractions, and PDFs are treated to be universal in nature. Partons can interact with each other and produce final state SM particles like leptons, and other partons as well. These partons are assumed to have negligible response to non-perturbative QCD effects, and can be treated as seeds of hadronic energy flows, termed as *jets* [19]. The experimental construction of jets will be discussed in subsequent chapters.

According to the above Lagrangian, gluons interact with quarks and anti-quarks and also with other gluons. However, since there are several generators and structure constants for the gauge group, it can be complicated to evaluate the exact colour charges. The Casimir invariants [13] of the fundamental and adjoint representations are C_F =4/3 and C_A = 3 respectively, which shows that the colour charge of a gluon is higher than that of a quark.

1.2.1 Strong coupling and asymptotic freedom

In pQCD, the predictions for observables are based in terms of $\alpha_s(\mu_R^2)$, where μ_R is a renormalisation scale. Considering $\mu_R \approx Q$ (energy scale) of a certain process, α_s can give an indication of the effective strength of the strong interaction in that process [20]. The dependence of α_s on the energy scale is termed as *running of the coupling*, and it can be described using renormalisation groups [21]. The running is logarithmic with Q, and satisfies the following renormalisation group equations (RGE):

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = \beta(\alpha_s) \tag{1.26}$$

where the beta function drives the Q dependence, and can be defined as,

$$\beta(\alpha_s) = -\alpha_s^2 (b_o + b_1 \alpha_s + b_2 \alpha_s^2 + ...) \tag{1.27}$$

with b_0, b_1, b_2 being the leading order (1-loop, LO), next-to-leading order (2-loop, NLO), next-to-next-to-leading order (3-loop, NNLO) coefficients respectively. The value of the strong coupling is conventionally specified using the value of Q^2 to be equal to M_Z^2 [22], and that leads to $\alpha_s(M_Z)$ approximated to 0.12.

It is seen that while describing QCD processes at other energy scales, a major portion of the quantum corrections can be absorbed into the running coupling. In particular, at higher

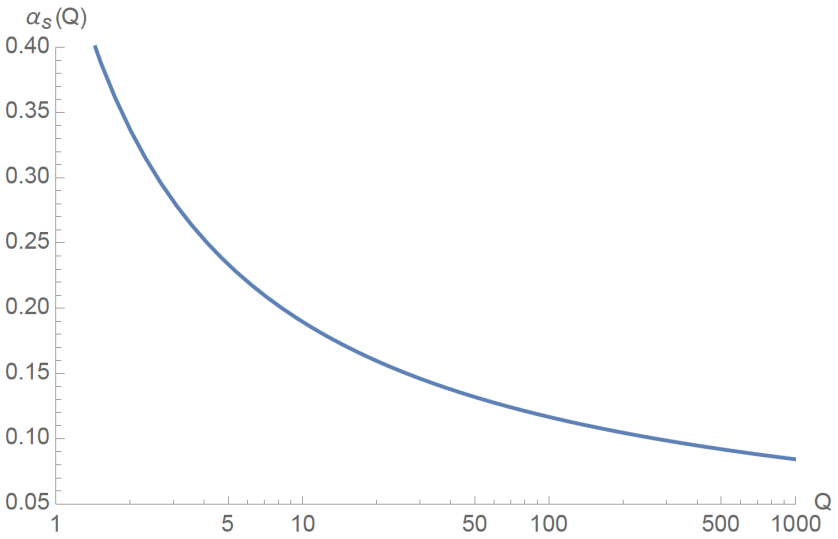


Figure 1.3: Running of α_s at LO, starting with $\alpha_s(M_Z) \approx 0.12$ [23].

momentum transfers (or energies), the quantum corrections can be described by a smaller coupling constant, as can be seen in Fig 1.3, and this behaviour is termed as *asymptotic freedom* [24, 25]. It can be described as follows,

$$\alpha_s(\mu) = \frac{1}{\beta_0 \frac{\mu^2}{\Lambda^2_{QCD}}} \qquad \beta_0 \approx 0.5|_{N_f=5} \tag{1.28}$$

where Λ_{QCD} is the QCD confinement scale [26] with mass dimensions and N_f are the number of active quark flavours, i.e. quarks having mass less than the scale μ. Asymptotic freedom is a pre-requisite for the success of pQCD, however, it should be noted that although α_s is small, it is not negligible, hence often higher order QCD corrections are necessary to obtain higher precision.

The knowledge of how scattering amplitudes behave in the soft and collinear limit is important for using pQCD to describe the hard scattering processes at the LHC. The soft limit is a scenario where the energy of an emitted gluon becomes small, whereas the collinear limit is the scenario where at least two particles propagate is the same direction, thereby having a small relative angle between their momenta. Scattering amplitudes tend to infinity when the soft or collinear limit is considered. Infrared and collinear (IRC) dynamics as shown in Fig 1.4 is non-perturbative [19], and cannot be used in an expansion in α_s terms, thus, understanding IRC limits of scattering amplitudes is necessary in order to construct observables that can be described using pQCD. Furthermore, soft and collinear emissions dominate high-multiplicity final states, and that is necessary to describe the evolution from hard-scattering at short distances to situations where QCD partons non-perturbatively transition to observable hadrons at large distances.

1.2.2 A non-arduous tour of the fundamentals of QCD inspired event generation

The matrix element (ME) level cross-section of a hard scattering process can be calculated using a Feynman diagram for the process of interest. Similar to the coupling, masses in QCD run as well due to quantum effects, and can also be characterised using μ_R. If one were to

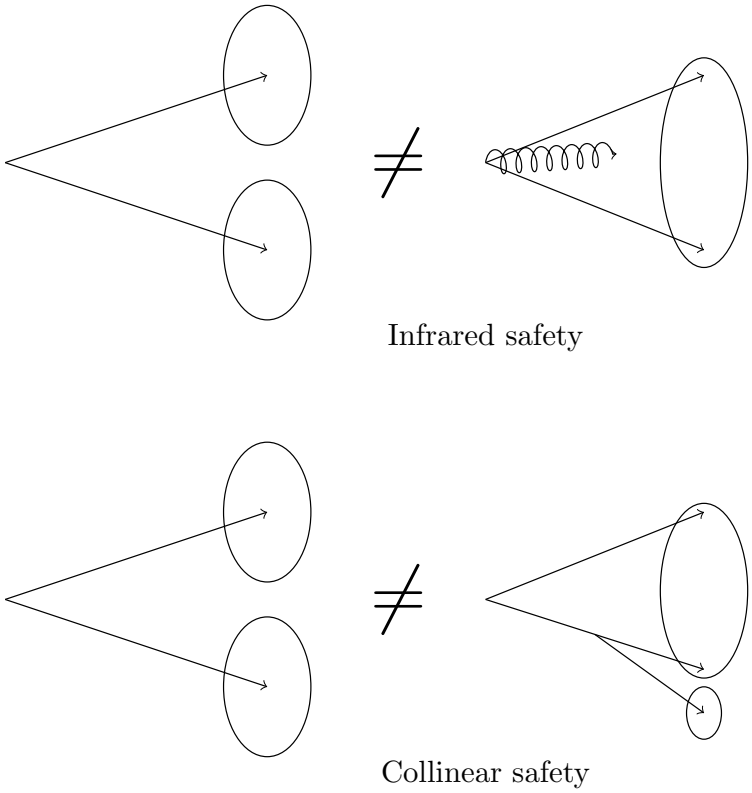


Figure 1.4: Schematic display of IRC safety. A soft gluon emission (top) between two jets should not result in one merged jet (infrared safety). A soft gluon emission at a small angle (bottom) from a jet should not result in changing the jet configuration (collinear safety) [27].

calculate an observable to all orders of perturbation theory, the dependence on μ_R would be cancelled. However the assumption of any truncation scheme leads to the possibility of a finite dependence on μ_R, which is related to higher orders in α_s [13]. This dependence on μ_R is usually quantified as a *theoretical uncertainty* by varying μ_R over a range. Considering $\mu_R^2 \approx Q^2$ of a certain process, the range can be defined as,

$$\frac{Q}{2} \leq \mu_R \leq 2Q \tag{1.29}$$

For most processes the effect of higher order variations are typically absorbed within this range. The energy fraction carried by partons entering a hard process is not known because of quantum mechanics, and hence a PDF becomes necessary. The longitudinal momentum fraction of the proton momentum carried by the parton, $x = p_{\text{parton}}/p_{\text{proton}}$, termed Bjorken-$x$, is defined as $f(x, \mu_F)$, where μ_F is the factorisation scale appropriate for the interaction. Thus, the initial state non-perturbative divergences can be absorbed in the factorisation scale. Again, as for μ_R, a finite dependence on μ_F is seen to arise, and can also be quantified as a theoretical uncertainty, varying μ_F similiar to 1.29.

The PDF evaluation is usually factorised into two parts. The non-perturbative part is parametrised by fitting to experimental data obtained from deep inelastic scattering experiments [28, 29], where electrons scatter off a proton, lose energy and get deflected, leading to the breaking up of the proton.

The perturbative component is obtained using Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [30–32],

$$\mu_F^2 \frac{\partial f_i(x_i, \mu_F^2)}{\partial \mu_F^2} = \sum_j \int_{x_i}^1 \frac{dz}{z} P_{ij}(z) f_j \left(\frac{x_i}{z}, \mu_F^2 \right) \tag{1.30}$$

which shows that the change of PDFs with the factorisation scale is easily calculable in perturbation theory. In order to make perturbative corrections small, it is sensible to associate the factorisation scale with a hard energy or momentum scale in the process. Thus, choosing $\mu_F \sim Q$, the DGLAP equations can point to the variation of PDFs with the energy scale.

To obtain the hadronic cross-sections, the partonic cross-section has to be combined with the PDFs, after setting the renormalisation and factorisation scales. However, even the hadronic cross-sections are far from what is observed in a hadron collider, since most processes in a collider are modelled as $2 \to N$ processes, with not all N of the final state (FS) particles being created at the hard scatter stage. Furthermore, the order of perturbation theory that is used to calculate the cross-section might not be sufficient for some of the measured observables. The *parton shower* (PS) [33–35] approach is taken in this scenario, i.e. starting from the $2 \to 2$ process, which defines the energies and directions of the hardest partons and then successively building up the full structure of the event with additional parton branchings. Often it is convenient to apply the PS approach only for leading logarithmics (LL)[3].

Additionally, a *k-factor* is used to correct for the cross-section of the hard process, if a NLO sample is not available. It is calculated using the ratio of the theoretically obtained (ME level) NLO or NNLO cross-section over the LO cross section. This *k-factor* is then used to reweight the event, without having to regenerate the event altogether. This does pose a problem that the difference of kinematics due to higher-order effects is not taken into account, but it is in general a widely accepted method for faster computations.

The branching processes $g \to gg$, $g \to q\bar{q}$ and $q \to qg$ are governed by QCD and determine the DGLAP equations, and the probability of a branching occurrence given by Sudakov form factors [36, 37]. Furthermore, the parton branchings can occur before and after the QCD interaction vertex leads to initial-state radiation (ISR) and final-state radiation (FSR). FSR is usually mass, p_{T} or emission angle ordered and termed as a *time-like* shower [38], whereas ISR is termed as *space-like* shower which involves backwards evolution [39], as shown in Fig 1.5. PS approach also assures IRC safety effectively till LL accuracy using the *resummation* approach [40], whereby the IR divergences due to higher-order perturbative contributions from virtual gluons are cancelled by the radiation of undetected real gluons.

Electroweak processes have non-negligible contributions at the LHC even though the dominant processes originate from strong interactions, and hence QED effects have to be properly estimated when simulating SM background processes. Since the $\mathcal{O}(\alpha_{EW}) \approx \mathcal{O}(\alpha_{\mathrm{s}})$, it is assumed that NLO EW effects are of similar magnitude compared to NNLO QCD effects, however, in certain phase space regions, their contributions can be non-negligible, i.e. virtual exchange of soft/collinear weak gauge bosons or photon emission.

[3]While identifying the jet by the initiating parton is most intuitive, it ignores the effect of additional radiation. The definition of born-level process is similarly ambiguous, and tying the definition into parton shower restricts us to LL accuracy.

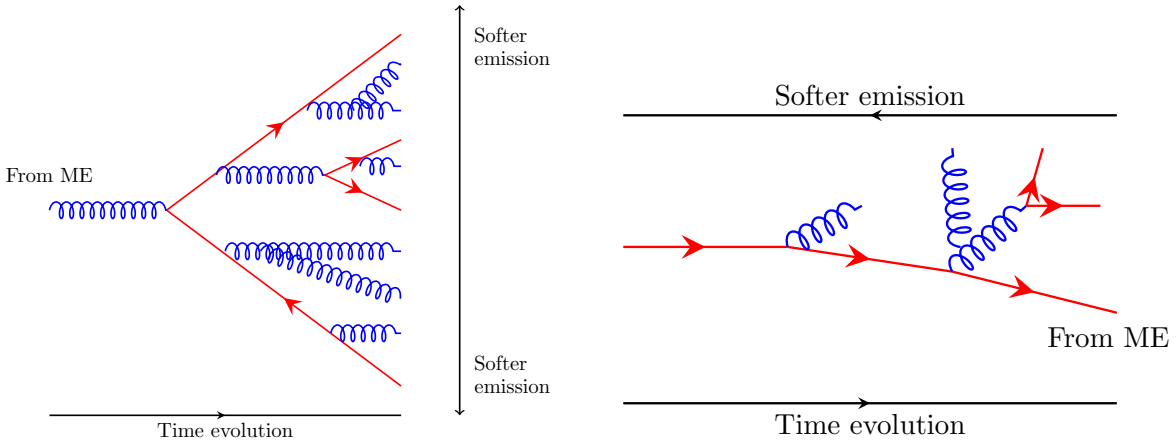


Figure 1.5: Schematic diagram of FSR (left) and ISR (right) shower development [27].

1.3 Monte Carlo event generators

Monte Carlo event generators (MCEG) like HERWIG [41], PYTHIA [42, 43], SHERPA [44] and POWHEG [45] perform detailed hard process calculations of cross sections assuming the underlying theory is the SM or one of its possible extensions. MCEG can model the physics of hadron collisions starting from the short distance scales, up to the usual hadron formation and decay scales, with the short distance physics being primarily based on pQCD convolved with parton distribution functions, since QCD is weakly interacting at distances below a femtometer. There are four major aspects that go into designing a MCEG, namely,

- perturbative computation of the primary process of interest, with decays of short-lived particles. Events after this stage are usually termed as parton level
- generation of QED and QCD radiation using PS
- non perturbative transition from partons to hadrons, i.e. hadronisation. Events after this stage are usually termed as particle level
- approximations for soft hadron physics and low-p_{T} interactions, i.e. generation of underlying event.

One can select a specific hard subprocess of interest at LO and partonic events are generated according to their matrix elements and phase space. Since the particles entering the hard subprocess, and some of them leaving it are usually partons, gluons are radiated, which can further radiate additional gluons or a quark and anti-quark pair, creating a partonic shower. Moreover, the composite nature of protons implies that multiple parton pair interactions (MPI) [46] can occur during the collision of two protons and that has to be accounted for when simulating the SM background processes. Each MPI is associated with a set of ISR and FSR showers, hence in the *interleaved* showering, MPI, ISR and FSR compete, with the highest ordering variable being allowed to determine the evolution [47, 48]. PS starts at the hard scale and evolves downwards to energy scales near the QCD confinement scale (Λ_{QCD}. This implies a decrease in the energy scale at every new branching and the evolution continues until the parton energy scale is of the order of Λ_{QCD} and they can transition to colour neutral hadrons, by the process of fragmentation [49] and hadronisation [50, 51]. The hadronisation process ideally occurs out of the proton radius and is non-perturbative in nature. There are two approaches when it comes to hadronisation as shown in Fig 1.6, namely,

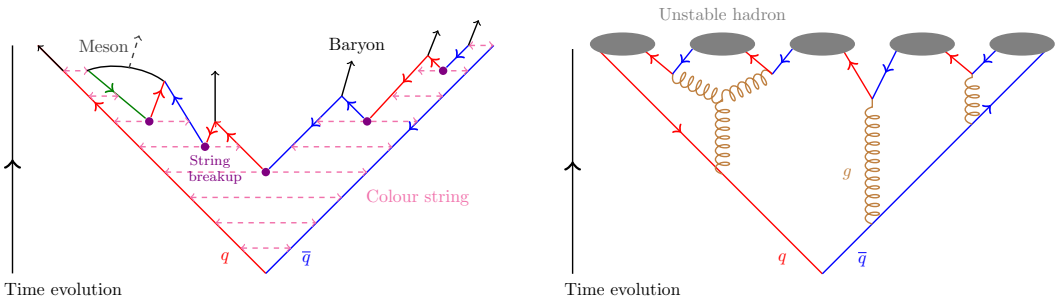


Figure 1.6: Schematic diagrams of Lund string (left) and cluster (right) hadronisation steps [27].

- Lund string model [52, 53] - Based on the idea that the potential between a q and $\bar{q}$ at short distances is alike to the electromagnetism potential, i.e. is inversely proportional to the distance. However, as the distance increases the behaviour changes, since unlike the photon, gluons are self-interacting. The QCD potential is linear in nature as shown in Fig 1.7, and each $q\bar{q}$ pair is assumed to be connected by a massless relativistic string, having no transverse degrees of freedom. The potential energy stored in the string increases as the pair moves away from the production vertex and can break to form new $q'\bar{q}'$ pairs, forming two colour-singlet systems of $q\bar{q}'$ and $q'\bar{q}$. The gluon field between the quark charges are kinks on the system. The breaking continues as long as there is enough energy to form hadrons.
- Cluster model [34, 49, 54] - Based on the property of *preconfinement* [55], whereby after a PS it is possible to form colour-singlet clusters of radiated partons and the mass distribution of these clusters are calculable from first principles. A standard approach is to split any remainder gluons after the shower into $q\bar{q}$ or diquark pairs. In the large-N_c approximation [4] [56], it can be shown that the adjacent di-/anti-quarks will have matching (anti-)colours, and can be associated with a single colour-singlet cluster. After the clusters are formed, they are matched to a smooth hadron mass spectrum.

Additionally, partons coming from hard process, MPI and beam remnants can interact with each other by exchanging colour, depicted as a connection of colour strings between them, and these interactions can lead to merging of two strings, termed as colour reconnection [57, 58]. Colour reconnection reduces the string length, and hence the multiplicity of the resultant hadrons. The hadronisation of a coloured system has more free parameters, owing to the fact that individual partons do not hadronise independently, but rather collectively if they are colour-connected. Hence, *tuning* [5] the model to one particular dataset does not restrict its predictive power for other collision types or energies.

All the particles in a scattering event which is not originating from hard scatter, and consists primarily of beam-beam remnants (BBR) and MPI is grouped together and broadly

[4] Despite the fact that QCD is an SU(3) gauge theory, it is frequently useful to assume a generalisation to a theory with N_c colours, SU(N_c). For any N_c, a fundamental colour can be combined with a fundamental anti-colour to form an adjoint colour and a colour-singlet. In simpler terms, the colour of a gluon can be thought of as being that of a quark and an anti-quark, with corrections owing to the fact that gluons do not have colour-singlet components. It is important to note that on using the large-N_c limit, corrections to it are expected to be suppressed by $1/N_c^2 \sim 10\%$, however depending on the topology in question the suppression factor dynamically varies.

[5] The process of determining parameter values using data distributions sensitive to them, in scenarios where the values cannot be obtained from first principles.

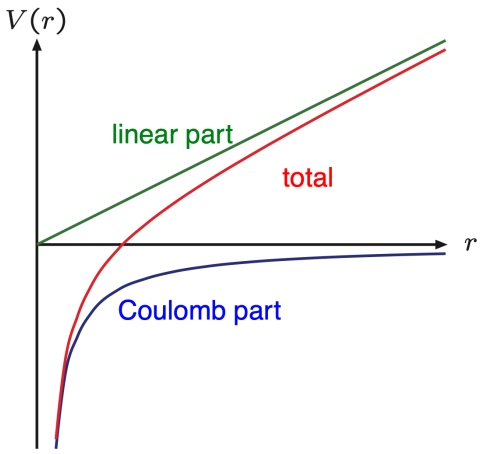


Figure 1.7: Potential between a $q\bar{q}$ pair, as a function of distance between them.

termed as *underlying event*, although there can be subtle dependencies of this physics with the actual process of interest. It is important to account for this "background" phenomena in order to precisely model the hard process behaviour and probe physics beyond SM.

In general, ME generators can do multi-loop (only automated at one-loop level) or multi-leg calculations whereas PS generators are accurate till LL and deal with the full event evolution. NLO calculations are usually more robust in terms of scale choices, since there can be kinematic features that seen when calculated at NLO, and hence it makes sense to combine the two to get a more precise result for the process under consideration. However, running them back to back can lead to double counting of the resultant jets, since the PS algorithm will take into account multi-legs/loops independently of that done by the higher-order ME calculation. This situation can be well described following the example of a Z+jets production, where the additional jet can come from NLO calculation at ME level, or from QCD radiation at PS level. In order to avoid this double counting, and get a precise result, it is useful to apply a *matching/merging* scheme [59, 60], that essentially decides a cut-off scale (based on p_{T} or energy of jets) above which ME is used, and below which PS is used. It should be noted that at LO+PS level, matching can also be an issue, where the double counting arises between the extra emissions from the hard process and the PS. There are several schemes available for this purpose, such as,

- CKKW (Catani-Krauss-Kuhn-Webber) [61, 62]: PS is started from the highest possible scale, and if there are shower emissions above the matching scale, they are vetoed. Each event is reweighted based on Sudakov form factors and running coupling from PS to avoid dependence on the matching scale.

- MLM (Michaelangelo Luigi Mangano) [63]: Original ME partons are matched to the jets obtained from PS[6] and if the event has all the jets from PS matched to a ME parton, it is retained. However, this approach throws away a lot of events, making the event generation process inefficient.

- CKKW-L (CKKW-Lonneblad) [64, 65]: Basically CKKW, but the weight calculation is done based on PS histories.

- NLO merging: Using POWHEG approach [66], momentum of the hardest emission is treated as a cut-off threshold to avoid double counting. Alternatively in MC@NLO ap-

[6] usually exploiting the distance parameter discussed in detail in the next chapter

proach [67], negative weights are assigned for PS configurations which potentially contain radiations already considered in the NLO calculation, and ultimately the weights are summed to obtain the correct normalisation of the sample.

A full event generation is depicted schematically in Fig 1.8.

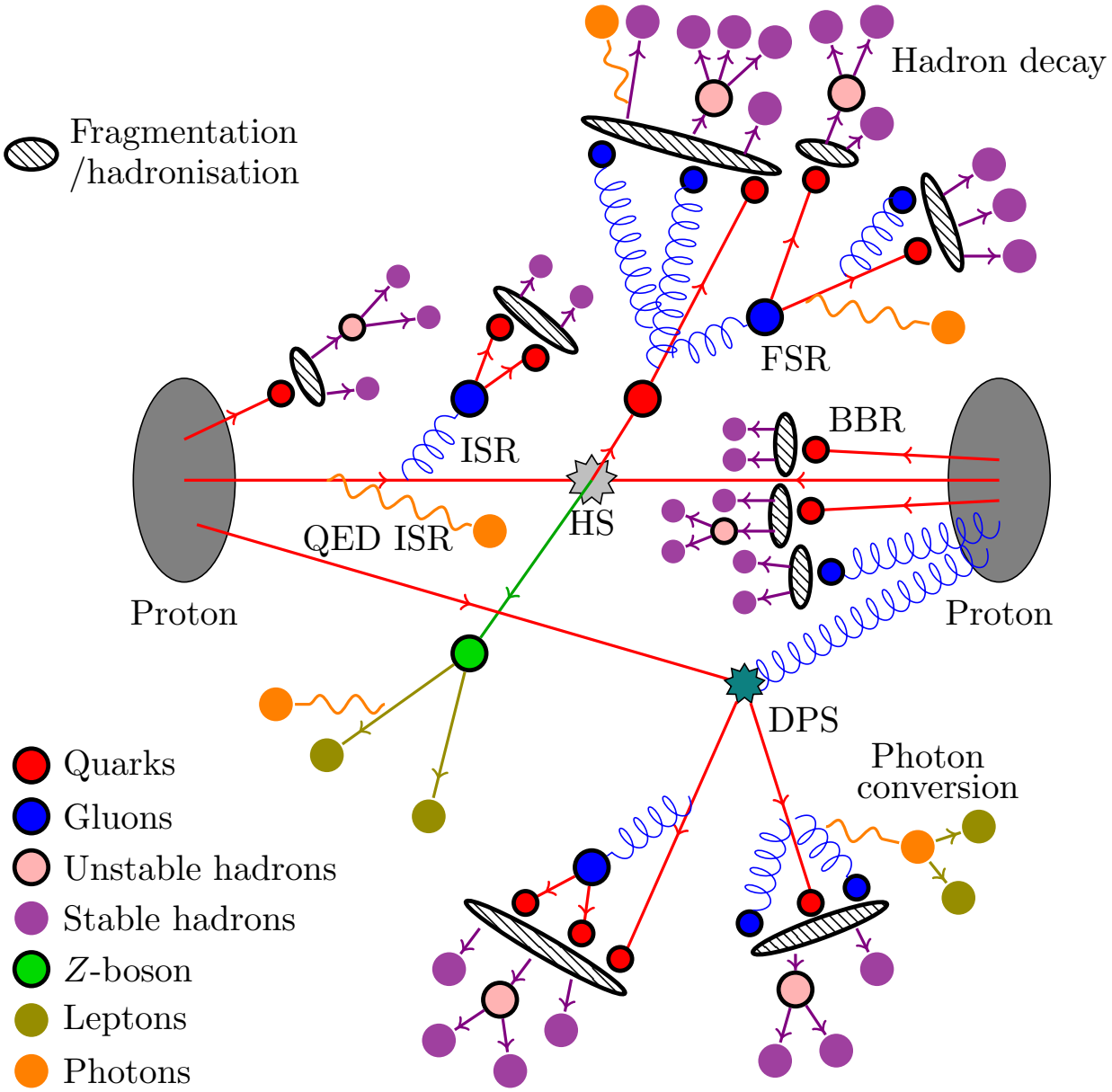


Figure 1.8: Simplified schematic diagram of a complete event generation, with a Z+jets process. The hard scatter (grey star) results in a Z boson and a quark. The Z boson decays to oppositely charged same flavour lepton pair, and one of those leptons emit a photon. The quark creates a PS via FSR, resulting in the formation of unstable hadrons which subsequently decay to stable hadrons. There is also QED and QCD ISR which results in a photon, and an independent PS respectively. A double parton scattering (DPS) results in two quarks each creating its own PS, with a gluon also radiating a photon that leads to two leptons. BBR also produces stable hadrons [27].

MCEG are used extensively in collider data analysis, to simulate standard model background and new physics signal events. These events are used to calibrate the physics objects

based on the simulation of detector response (as discussed in Chapter 3) or to devise ways to discriminate signal against background and design analysis strategies (as discussed in Chapter 4). It is of critical importance to classify particles from MCEGs based on their origin in analysis setups in a generator-agonistic way, as many MCEGs are used in ATLAS and they have different ways of presenting events records. A theoretically well-motivated way to classify particles originating from MCEG was formulated as a part of this thesis, presented in Chapter 6, which is now the default in ATLAS.

1.4 The limitations of SM

While the SM provides a very solid ground for understanding how nature behaves, there are still a few open questions which cannot be addressed by the theoretical framework. Although the discovery of the Higgs boson [68, 69] completed the SM, there lies an unexplained large discrepancy between the electroweak scale ($\mathcal{O}(100$ GeV)) and that of the Planck scale ($\mathcal{O}(10^{19}$ GeV)), and this is termed as the *hierarchy problem* [70]. From the large Higgs propagator correction, it is expected that the Higgs mass is sensitive to mass scales beyond the SM, e.g. the Planck scale. However that is not the case, so the question remains if nature is really so precise and finely-tuned, or there is a possibility of new physics that stabilises the Higgs boson mass to be approximately 125 GeV, by protecting the Higgs mass from arbitrarily high scales.

Traditionally SM treats neutrinos as massless, however, experimentally it is confirmed that neutrinos do have a small non-zero mass. This evidence arises from the observation that a neutrino of a particular type or "flavour", i.e. a muon neutrino can transform into a neutrino of another flavour, i.e. a τ neutrino. This process of transformation is termed as *neutrino oscillations*, and implies that neutrinos possess mass [71]. It is natural to assume that like the different generations of quarks and charged leptons, neutrinos have an ascending mass hierarchy through the generations, however there is no conclusive evidence of that scenario, and this behaviour of neutrinos are not accounted for in the SM.

Despite the fact that SM creates matter and anti-matter in pairs, there is a visible abundance of matter over anti-matter in the observed universe, with no appropriate explanation, and this open problem can only be explained by the requirement of a baryogenesis mechanism [72]. The strong-CP problem [73] is yet another flaw of the SM. It ties to the question as to why there is no CP violation in strong interactions. The strong sector naturally allows for CP violation, however it has not yet been experimentally observed.

However, the most important drawback of SM which is the focus of this thesis, is its inability to account for dark matter (DM), which is known to comprise 27% of the energy budget of the universe. There is astrophysical evidence that hint at the existence of DM and they will be discussed in length in the next chapter. Up until the 1980's, SM neutrinos were thought to be a potential DM candidate, due to their non-luminous nature, but the idea was dropped once it was experimentally found that neutrinos had the potential of forming superclusters initially and would then break down to form galaxies, which is contrary to observations [74]. The nature of DM is elusive – it could be yet another particle, and that hints at the possibility that the SM might in fact not be a complete theory of the observable universe.

Chapter 2

Strongly interacting dark sector - why and how?

In the previous chapter we discussed the basic idea of SM and alluded to why it might not be a complete theory. This chapter will touch upon the idea of dark matter (DM) and the possible candidates of particle DM. However, there is no reason to believe that the particle nature of DM is manifested through the existence of just one additional particle. It can be a plethora of particles in a *dark sector* (DS), much like its SM counterpart. The focus of the second half of this chapter will be on such a class of models which try to incorporate a strongly interacting DS of particles, building upon the standard QCD knowledge.

2.1 Moving on to... the overview of dark matter

Over the years there have been several theories postulating the nature of DM, with records dating back to the early 1930s, which states that the universe is comprised of an unknown form of non-baryonic matter that influences the structure of the cosmos. In the following decades, astrophysical and cosmological studies, including the study of cosmic microwave background (CMB) radiation, have indicated indirect evidences for the existence of DM, some of which will be discussed in this section. Yet, till date, no direct evidence of the particle nature of DM has been found at the Large Hadron Collider (LHC). However, in order to probe the nature of DM, it is necessary to appreciate the different ways in which DM can potentially manifest itself in the observable universe.

2.1.1 Early explorations

The pioneering work of astronomer F. Zwicky in 1933 [75] – studying the redshifts of Coma galaxy clusters – pointed out that there is a scatter in the apparent velocities of the eight galaxies within that cluster. These large velocity dispersions meant that the cluster density was higher compared to the one that can be obtained only by visible matter, and in turn led to the pre-emptive prediction of the existence of a non- luminous matter, with much higher density than radiating matter [76, 77].

This initial work triggered a wave of experimental and theoretical studies by different groups of astronomers. Although all of the studies cannot be discussed in great detail, as it is beyond the scope of this thesis, it is important to point out the major landmarks that govern our current understanding of DM. One such group of studies pertains to the *rotation curves of galaxies*, which is the circular velocity profile of the gas and stars in a galaxy, with respect to their distance from the galactic center. In 1970s, V. Rubin *et al.* [78] and

A. Bosma [79] studied the rotation curves of spiral galaxies using optical images since it is possible to infer the mass distribution of galaxies from their rotation curves. They found almost flat rotation curves, with no decline in the outer reaches of the galaxies, which was in contradiction to the expectation, if only visible matter existed. By early 1980s, substantial work went into measuring the rotation curves of galaxies well beyond their optical radii using 21 cm line of neutral hydrogen gas [80, 81], and confirmed the existence of DM.

On the other hand, experimental explorations of *cosmic microwave background* (CMB) [82, 83] – a uniform background of microwaves surrounding all directions in space – established the modern paradigm of the hot big bang cosmology. CMB radiation is estimated to have originated 3.8×10^5 years after the big bang, when the universe was filled with hot, ionised gas. Subsequently, free electrons combined with protons to form neutral hydrogen atoms (termed as *recombination*). Before this epoch, the CMB photons were tightly coupled to the baryons. Recombination increased the range of CMB photons from short length scales to long length scales, and led to their decoupling, and freely travelling through space. The CMB photons thus retain information about the state of the universe during the recombination timescale, and consequently, about the general nature of matter in the early universe. The observed small fluctuations in CMB increased with time due to the presence of gravity and ultimately lead to the formation of galaxy clusters that are currently observed. The shape of the amount of fluctuations in the CMB temperature spectrum at different angular scales – termed as the *power spectrum* – is determined by the primordial fluctuations of the inflaton [84] along with hot gas oscillations in the early universe. The amplitudes and resonant frequencies of these oscillations depend on the composition. Analysing this power spectrum can provide a strong handle towards mapping the density of matter in the early universe, and the latest studies report that the energy budget of the universe is composed of ∼27% DM, ∼5% visible matter and ∼68% Dark Energy [85].

Although there are several sources of indication for the existence of DM, they are merely indirect evidences through the gravitational influence of DM on visible, baryonic matter. However, the question still remains as to what is the nature of DM? It is certainly non-baryonic since the measured abundance of light elements that was produced in the primordial nucleosynthesis is much smaller than the total density of matter. One plausible explanation could be that DM is yet another individual (set of) elementary particle(s), not yet discovered. Furthermore, since visible matter comprises of several particles, there is no obvious indication that the same should not be expected for the *dark sector*. It is also well-established from astrophysical experiments that DM cannot possibly have EM charge, and in order for it to still be around, DM has to be long-lived compared to the age of our universe.

2.1.2 Dark matter relic density

Before diving into further specific discussions of DM, it is imperative to understand the abundance of DM that is expected to remain at this point of the lifetime of our universe, termed as *relic density*. It is believed that the generation of DM in the early universe can proceed via thermal or non-thermal production, or from a particle-antiparticle asymmetry.

There are predominantly two approaches when it comes to understanding the "particle" nature of DM.

- Following the standard cosmological model (ΛCDM) [86, 87], that is successful in explaining observations for the CMB, thermal history, and large-scale structure of the universe, DM is considered to be a cold, i.e. non relativistic particle (cold dark matter). In this scenario, DM is thought to be a pressure-less component of matter which was decoupled from the thermal bath (*freeze-out*) well before recombination timescale.

There is an alternative approach, termed thermal (*freeze-in*) involving a Feebly Interacting Massive Particle (FIMP) that interacts so feebly with the thermal bath that it never attains thermal equilibrium. The relic density in this case is a combination of initial thermal distributions together with particle masses and couplings, measured in a controlled experiment or astrophysically.

- It can also be assumed that larger halos (superclusters) would form first and then later fragment into smaller halos (galaxies), and the only way that would be possible is if the DM component is considered to be relativistic (i.e. hot) during the beginning of the universe.

The former idea is particularly favoured, since there is no reason not to assume that like CMB photons, neutrons and other light elements, DM also originated from a thermal decoupling process. The latter idea is usually disfavoured since hot DM would lead to a very different galactic structure, which is not in tune with current astrophysical observations.

2.1.3 Dark matter particle candidates

Cold dark matter is thus considered to be a stable nature. It is long-lived enough to form galaxies first and eventually form superclusters, and the potential candidate for this type of dark matter are the weakly-interacting massive particles (WIMPs) [88]. The idea of WIMPs arises in several theories that aim to resolve other limitations of the SM. The supersymmetric extension of SM (SUSY) [89] which is essentially an additional symmetry between fermions and bosons, that allows for inter-conversion of fermions and bosons, deserves specific mention in this context. According to the SUSY hypothesis, every fermion is said to have an associated superpartner boson and vice versa, effectively doubling the number of particles. Under the SUSY scheme of particles, there are several potential electrically neutral and weakly interacting dark matter candidates, like the neutralino (superposition of neutral partners of Higgs and other gauge bosons), the gravitino (superpartner of graviton, coming from the quantum theory of gravity) and the sneutrino (superpartner of neutrino). All these candidates may be placed under the general class of WIMPs and a comprehensive description of SUSY can be found at Ref [90].

Another possible class of particle candidates of DM are the axions, or axion-like particles (ALPs). As discussed briefly in the previous section, axions were originally introduced as a solution to the strong-CP problem, by virtue of an additional global symmetry of the theory which is spontaneously broken below an energy scale, f_a [73, 91]. The axion mass is seen to be inversely proportional to f_a and can solve the strong-CP problem for a wide range of values. Apart from this purpose, axions can also be treated as candidates of *cold dark matter* since they are assumed to be produced in the early universe non-thermally. Similarly, ALPs [92] arise in the spontaneous breaking of a global symmetry, however in this case there is no strict relation between the mass of the ALPs and the energy scale f_a, and they can be treated as viable candidates for DM.

2.1.4 Interactions between Standard Model and dark matter

The different types of DM searches that can be probed when the WIMP scenario is assumed, are shown in Fig 2.1.

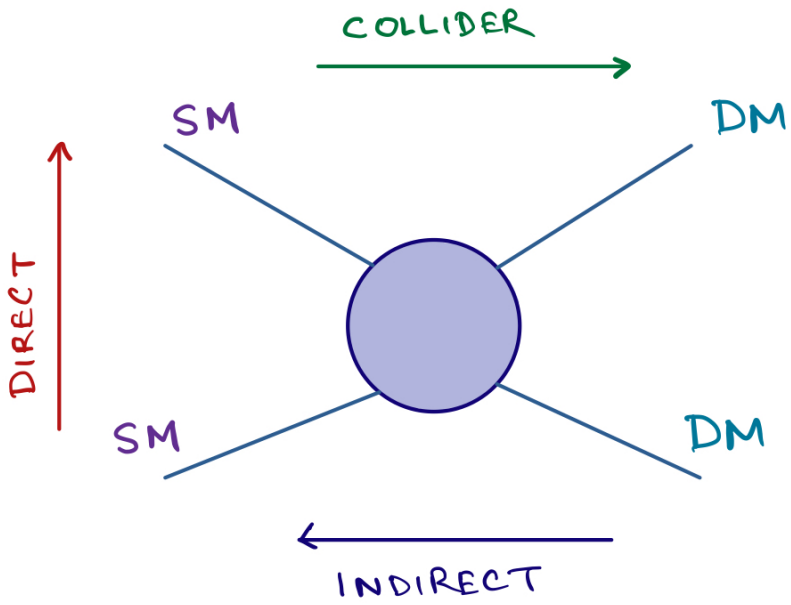


Figure 2.1: Types of DM searches governing the WIMP paradigm. Annihilation of DM particles and observing the SM outgoing particles is the indirect detection approach, the scattering of a DM particle off of a SM detector is the direct detection approach, and the annihilation of SM particles to form DM particles is the collider approach.

Indirect detection (non-collider)

Indirect detection methods focus on looking for products of DM interactions, particularly leading to SM particles, rather than the DM itself. An annihilation cross-section of the order of 10^{-26} cm^{3}s^{-1} is expected from the measured cosmological DM density, and hence indirect detection methods try to probe the outcome of the annihilation of DM particles. If DM is assumed to be unstable, it would decay and produce observable decay products, and there are several experiments which have been devised to detect DM decay products using γ-rays e.g., VERITAS [93], HESS [94], MAGIC [95], FermiLAT [96], X-rays e.g., XMM-Newton [97], NuSTAR [98], Suzaku [99], neutrinos e.g., IceCube [100], ANTARES [101], Baikal [102], Baksan [103], Super-Kamiokande [104], cosmic rays e.g., HAWC [105] and detectors in space e.g., DAMPE [106], CALET [107], AMS [108].

Despite the fact that none of the above mentioned experiments have been able to report any signal so far, they have been successful in narrowing down the parameter space for future indirect searches. Several of these experiments are already undergoing upgrades (IceCube-Gen2 [109], KM3NET [110], CTA [111], Baikal-GVD [112], Hyper-Kamiokande [113]) that would set the stage for higher sensitivity of indirect DM searches. A very detailed review of this aspect of DM searches can be found at [114].

Direct detection

Direct detection of dark matter involves low energy (sub-MeV) scale scattering events, where the main focus is to try and record the rare events when a DM particle scatters off a target material, with negligible background interactions. The driving idea is that dark matter was produced thermally in the early universe and subsequently decreased via annihilations into ordinary matter, to attain a stable equilibrium. The primary signal of DM scattering

is nuclear recoils from Earth-based detectors, and direct detection of WIMPs is a widely recognised research area, originally proposed in [115]. Since WIMPs have no electric charge, they should mostly scatter off the atomic nucleus, and subsequent momentum transfer gives rise to a potentially detectable nuclear recoil. The dominant part of nuclear recoil energy induced by WIMP-nucleon interaction is lost as heat, and this leads to atomic motion, which in turn gives rise to phonons [116] in solid materials. The available electronic energy loss can excite or ionise the target atoms, and this leads to narrow emission spectrum scintillations which can be detected by photosensors.

Several direct detection experiments like CRESST-III [117], XENON100 [118], DAMA or LIBRA [119] are searching for DM signal excesses over the background, but have not confirmed any positive results yet [120, 121].

Collider searches

When it comes to collider searches, they can broadly be classified into, searches for UV complete models, i.e. SUSY, or searches for simplified models. The search performed in the main body of the thesis uses a simplified model of DM as a benchmark, with the assumption that the mediator connecting SM and dark sector is beyond the scale of interaction, and results in a contact interaction. Since, DM is non-luminous in nature, hence the collider final state is expected to contain a significant amount of missing transverse momentum. But, the more important question to ask is, what do we measure, if everything is invisible?

The only SM particle which does not interact with any of the detector components[1], is the neutrino, since they are colour- and charge- neutral, and have weak interactions. Hence, neutrinos constitute missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$). This can be realised in accordance with the energy-momentum conservation principle which demands that the total p_{T} of all particles has to be zero after collision, since the initial particles (protons) move along the beam axis. Hence, $E_{\mathrm{T}}^{\mathrm{miss}}$ can be treated as the negative sum of scalar p_{T} of all other outgoing objects.

In order to make the search for dark matter in colliders relatively model-independent, it is assumed that the recoiling visible particles should be governed by Standard Model interactions. There are a significant number of models which predict the possibility of SM bosons being present in any beyond Standard Model (BSM) process, which contribute to initial state radiation (ISR) from partons, and since hadron colliders are dominated by gluon ISR, hence looking at *mono-X* ($X = \gamma, h, Z, top, jets$ etc) final states is one of the most used approaches in DM searches, and this will be further discussed in the next chapter.

For the context of this thesis, we will restrict ourselves to the alternative (non-WIMP) collider searches paradigm for DM. The remainder of the chapter will discuss the primary type of DM model that was probed for this thesis, in terms of the theoretical considerations and the model parameters.

2.2 Simplified Models of dark matter

Several new theories have been proposed where the DM exists within a new dark sector [122–126], defined by a new set of particles, forces and interactions, and this dark sector can interact with the SM sector via portal interactions, with the renormalisable examples being the lepton, photon and Higgs portals [127]. In scenarios where the BSM mediator is massive compared to the collision energy, the interaction between SM and DM can be modelled as a

[1] will be discussed in the next chapter

contact interaction, and effective field theories (EFTs) can be used to explain the production of invisible particles. In general, dark matter EFTs follow the simple principle that the DM is a single particle candidate beyond the SM and all the other degrees of freedom are either negligible to have any significant impact in the observable spectrum, or are heavy enough to be integrated out [128]. This is determined by one parameter (dimension-D operator) that controls the production rate, and the corresponding Lorentz structure is seen to have a mild effect on the kinematic distributions of the invisible particles. The interaction Lagrangian contains all the gauge and Lorentz invariant terms and the additional DM field (χ) term can be written as an expansion of dimension-D operators:

$$\mathcal{L}_\chi = \sum_{D,i,f} C_i^D \mathcal{O}_{i,f}^D \tag{2.1}$$

The Wilson Coefficients (C_i^D) [129] are parametrised as $C_i^D = \Lambda_i^{4-D}$, where Λ_i is the cut-off scale for the EFT. EFTs are a more generalised approach for studying the dark sector, since the dimension-D (D=6) operators can range over several mediation schemes, and the complicated high-energy interactions are integrated out in the UV completed contact operator limits.

In cases where the mediator mass is much less than the collision energy, simplistic descriptions of collider phenomenology can be constructed, not considering the extra physics details of the theory observed at energies greater than those accessible at the collider and irrelevant at the LHC. These descriptions are collectively called simplified models, and can be treated as a transition step between full theories and simplest EFTs. The primary idea of simplified models is based on the fact that only a small number of new particles may be present, and these models can be developed from tree-level interactions (pair production, associated production) of invisible particles, which is mostly sufficient for discussing LO collider phenomenology [130].

2.3 QCD-like dark sector

A non-Abelian dark sector which displays asymptotic freedom similar to SM QCD, can be categorised as a QCD-like DS, and it is assumed that in such scenarios, the SM sector and DS are coupled through a portal [131, 132]. The most general class of such models involve a minimal extension of the SM with new particles and couplings. It can be assumed that the new mediator acts as a portal.

In s-channel production mode shown in Fig 2.2, the mediator is a colour singlet and has to couple to a pair of either DM particles or SM particles. Since massive colour-neutral spin-1 bosons with axial-vector or vector couplings are present in many BSM theories, Z' bosons can be treated as prospective mediators in most models. Here a Z_2 discrete symmetry is assumed, under which the SM fields and the mediator are even, but the DM particle is odd, to guarantee DM stability [133–139].

In t-channel production mode (Fig 2.2), the mediator interacts with DM and one of the SM quarks, and usually simplified models consider the scenario of fermionic DM particle which interacts with SM particles via a scalar mediator coupling only to right-handed quarks. A generic t-channel DM simplified model contains an extension of the SM by two additional fields: a DM candidate, χ and a mediator, ϕ which has its fundamental representation in $SU(3)_c$ and a dark non-Abelian gauge group [140–143].

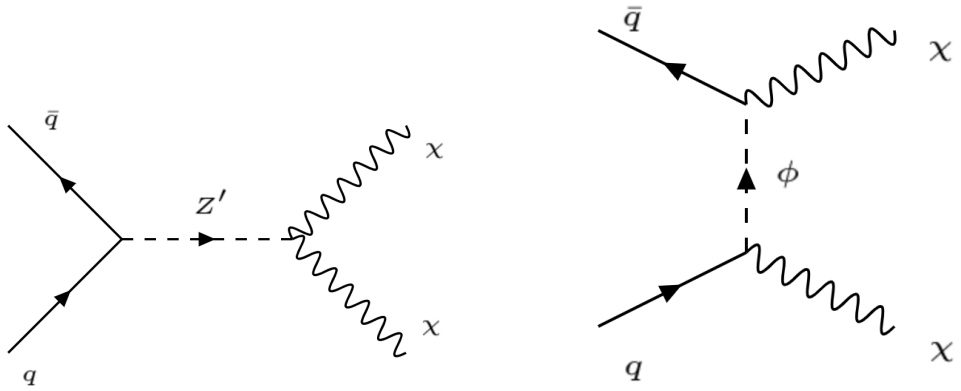


Figure 2.2: s-channel and t-channel production modes for DM production.

2.3.1 Semi-visible jets theory model

Semi-visible jets [123, 136, 143] (SVJ), are jet-like collider objects where the visible states in the shower are Standard Model hadrons. It is assumed in these scenarios that the strongly coupled hidden sector contains some families of dark quarks which bind into dark hadrons at energies lower than a dark-confinement scale Λ_d. Reference [143] targets a non-WIMP scenario which eventually leads to interesting collider signature, where the final state consists of a jet aligned with missing transverse momentum[2] ($E_{\mathrm{T}}^{\mathrm{miss}}$) due to a mixture of stable, invisible dark hadrons (with decay time $c\tau > 10$ mm) and visible hadrons from the unstable subset of dark hadrons that promptly decay back to SM particles. The total momentum of the dark matter is hence correlated with the momentum of the visible states, leading to event $E_{\mathrm{T}}^{\mathrm{miss}}$ close to a jet.

The model of Ref. [143] uses a simplified parameterisation (an extension of simplified models), where a direct mapping of the Lagrangian parameters to physical observables is not possible since some of the dark sector observables depend on non-perturbative physics. The three parameters of this model in the t-channel production mode are:

- Mass of the scalar bi-fundamental mediator (Φ), which is in the fundamental representation under both visible QCD and the dark non-Abelian gauge group (for t-channel production), denoted by M_ϕ.
- Dark hadron mass, denoted by M_{D}.
- The ratio of the number of stable dark hadrons over the total number of dark hadrons in the event, r_{inv}.

The first two parameters are set during the ME level event generation stage, whereas the third parameter is set during the dark shower stage of event generation, which helps to achieve the desired collider topology in a simple manner. r_{inv} in its intermediate regime makes the SVJ appear, by controlling the invisible energy fraction as shown in Fig 2.3, hence higher the value of r_{inv}, more $E_{\mathrm{T}}^{\mathrm{miss}}$ the event will have.

In order to introduce a simple portal, the BSM effects are parametrised in an EFT expansion. In this approach, an effective Lagrangian captures all possible interactions:

[2]The initial momentum of the colliding partons along the beam axis is unknown because the energy of each hadron is split and constantly exchanged between its constituents, so the amount of total missing energy cannot be determined. However, the initial energy of particles travelling transverse to the beam axis is zero, so any net momentum in the transverse direction indicates missing transverse momentum, $E_{\mathrm{T}}^{\mathrm{miss}}$. More details will be discussed in the following chapter.

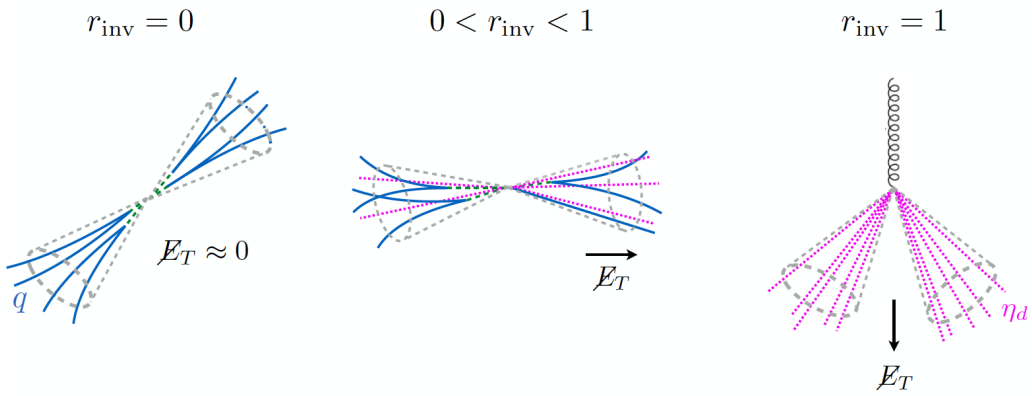


Figure 2.3: Diagram showing the direction of $E_{\text{T}}^{\text{miss}}$ for different r_{inv} values [143].

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + (1/\Lambda_d)\mathcal{L}_1 + (1/\Lambda_d^2)\mathcal{L}_2$$

where $\mathcal{L}_i$'s are constructed from Standard Model operators that obey the $SU(3)_C$ x $SU(2)_L$ x $U(1)$ gauge symmetries, and the higher-dimensional Lagrangian terms representing effective (i.e. non-fundamental) couplings, are suppressed by powers of Λ_d. As the dark matter particles can appear in the final state of this model, a dark matter Effective Field Theory approach is used, in which the DM is the only additional degree of freedom beyond the SM accessible by current experiments, and hence the interactions of the DM particle with SM particles are described by effective operators (of dimension-6 or higher) of the form:

$$\mathcal{L}_{\text{contact}} \supset (c_{ij\alpha\beta}/\Lambda^2)(\overline{\text{q}}_i\gamma^\mu q_j)(\overline{\chi}_\alpha\gamma_\mu\chi_\beta)$$

Here, q are the SM quarks, χ are the dark sector quarks, Λ is the scale of the operator, and $c_{ij\alpha\beta}$ are $\mathcal{O}(1)$ couplings containing the information of possible flavour structure. The Lagrangian containing the interaction and kinetic terms is:

$$\mathcal{L}_{\text{dark}} \supset -\tfrac{1}{2}\,\text{tr}\, G^d_{\mu\nu}G^{d\mu\nu} - \bar{\chi}_a\left(i\not{D} - M_{d,a}\right)\chi_a$$

Here, the dark sector is a $SU(2)_D$ gauge theory with coupling $\alpha_d = g_d^2/4\pi$, containing two fermionic states $\chi_a = \chi_{1,2}$, and c_{ijab} are O(1) couplings that encode the possible flavour structures, and (assuming minimal flavour-violation) light-flavour production channels dominate.

If dark mesons exist, their evolution and hadronisation process are currently not very constrained. They could decay promptly and result in a very SM QCD like jet structure, even though the original decaying particles are dark sector ones; they could result in semi-visible jets; or they could behave as completely detector-stable hadrons, in which case the final state is just the missing transverse momentum. Apart from the last case, which is more like a conventional BSM $E_{\text{T}}^{\text{miss}}$ signature, the modelling of these scenarios is a fairly unexplored area.

2.3.2 Hidden Valley Shower

Pythia8 [43] is a general purpose MCEG used for simulating high-energy collisions, starting from a few-body process and resulting in multi-hadronic final states. It includes libraries containing tree level calculations for different hard processes and initial- and final-state parton shower, multiple parton interactions, beam remnants, fragmentation and hadronisation, and particle decays, and is used to model the SM process outcome in a detector. The tool

works well for SM, but when models with BSM interpretations have to be studied, there is no guiding principle for the underlying hadronisation. Many BSM models contain sectors with new gauge groups and matter content, which may decouple from the SM at low energy limits.

The Hidden Valley (HV) [123, 144–146] module of PYTHIA8 was designed in order to study a strongly coupled dark sector. The module tends to simulate a reasonably generic framework for studying DS models, and HV being a light hidden sector, the associated particles may have masses as low as 10 GeV and the spectrum of the valley particles and their dynamics depends on the valley gauge group G_v, their spin and the number of particles contained in the theory, along with their group representations. The construction of the HV sector begins with the specification of the particle content. There are 12 particles which are charged under both the SM and HV symmetry groups, with each particle coupling flavour-diagonally with the corresponding state in SM, but has a fundamental representation in the HV colour symmetry as well. They are listed below:

Table 2.1: List of Hidden Valley particles

Hidden Valley particle	**Corresponding SM particle**
D_v	d quark
U_v	u quark
S_v	s quark
C_v	c quark
B_v	b quark
T_v	t quark
E_v	e lepton
ν_{Ev}	ν_e, electron neutrino
μ_v	μ lepton
ν_{μ_v}	ν_μ, muon neutrino
τ_v	τ lepton
ν_{τ_v}	ν_τ, tau neutrino

These particle states are collectively referred to as F_v. In addition to these states, it is assumed that the HV contains either of the following two:

- Abelian U(1) gauge symmetry, broken or unbroken leading to a γ_v which is the massless gauge boson of the HV U(1) group.
- Non-abelian SU(N) gauge symmetry, which is unbroken and leads to $N^2 - 1$ massless gauge boson of the HV SU(N) group g_v's and new massive matter particles q_v's in its fundamental representation.

There are three alternative production mechanisms of these particles:

- A massive Z' can act as a mediator for s-channel production modes:
  $q\bar{q} \to Z' \to q_v\bar{q_v}$
- There is also a possibility of kinetic mixing terms:
  $q\bar{q} \to \gamma \to \gamma_v \to q_v\bar{q_v}$
- The final general possibility is the presence of a F_v which is charged under both the HV and SM groups, and hence can have the following production scheme:
  $gg \to F_v\bar{F}_v$ and subsequently decay to $F_v \to fq_v$, where f is a SM fermion

For the broken U(1) symmetry, hidden valley photon acquires mass and then decays to regular photons that make pairs of SM particles. As for the SU(N) symmetry, the dark quark hadronises in hidden sector with full string fragmentation evolution, and produces up to eight q_v flavours and 64 $q_v q_v$ mesons. The masses are assumed to be degenerate, but the meson masses and decay mechanisms can be specified by the user, depending on the model. The radiation off the HV charged particles is allowed, and HV particles are not produced in ISR. Since, all the (anti-)particles F_v and q_v have one (negative) positive unit of HV charge, hence the HV radiation is similar to that of QCD, so the showering mechanism for the three radiations (QCD, QED and HV) are interleaved. The emissions are arranged in one common sequence of decreasing emission p_{T} scales and hence it is difficult to separate one kind of radiation from the other.

The HV particles with no SM couplings are invisible and their presence can only be detected by observing the amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ present in a particular event. There are two possibilities of simulating such a scenario where activity in the hidden sector seeps through to the visible sector and constitutes $E_{\mathrm{T}}^{\mathrm{miss}}$.

- If the U(1) symmetry is broken, then γ_v acquires mass, and subsequently decays back to SM particles either by a Z' state or some either mediator, since it has a small but non-zero mixing angle with the SM photon. However, the q_v remains absent in this scenario due to the lack of any U(1) charge.

- In case of the SU(N) symmetries, the gauge group remains unbroken leading to massless gauge bosons g_v and there is confinement of partons. In this scenario, after the hadronisation, the q_vs'and $\bar{q_v}$s'can be obtained which can either decay back to SM or remain stable, depending upon the mixing of the states. The q_vs' can exist as stable and invisible states if they are off-diagonal and flavour-charged, whereas diagonal ones can decay back to the SM and contribute to formation of visible hadrons.

The HV module of Pythia8 allows the existence of visible jets (where the final state is like a regular QCD jet, even though the initial constituents came from the hidden sector), invisible jets (where the HV particles are completely stable and only contribute to missing transverse energy), and semi-visible jets (where some of the dark hadrons decay back to SM whereas the others remain stable within collider timescales) [132]. Additionally, depending on the lifetime of the dark sector particles, i.e. when they are long-lived, they can lead to emerging jets [147]. The different possible final states that can arise from the HV module is shown in Fig 2.4.

While the detailed description of the HV module is beyond the scope of this thesis, it essentially builds a dark-sector parton shower emulating the QCD shower. The specific parameters choice in Pythia8 HV module have been listed in Table 2.2. The dark hadrons obtained in the Pythia8 HV showering, i.e. flavour-diagonal HV-meson with spin 0/1 that can decay back into the Standard-Model sector) have masses of 20 GeV, while the off-diagonal HV-meson with spin 0/1 that is stable and invisible have masses of 9.99 GeV. Studies discussed in [139, 148–150] have shown that the decay of dark hadrons also depends on the mediator to the visible sector. Two different dark quark flavours combine to form dark π^+, π^-, π^0 and ρ^+ , ρ^- , ρ^0, where the dark ρ are assumed to be produced thrice as much as pions. The dark $\rho_{d(\mathrm{unstable})}$ mesons tend to decay promptly via $\rho_{d(\mathrm{unstable})} \rightarrow \rho_{d(\mathrm{stable})}\rho_{d(\mathrm{stable})}$ decay channel, whereas the dark $\pi_{d(\mathrm{unstable})}$ decay promptly via $\pi_{d(\mathrm{unstable})} \rightarrow \pi_{d(\mathrm{stable})}\pi_{d(\mathrm{stable})}$ decay channel, except for the ρ_d^0 meson, which decays into SM particles, following a five flavour scheme, due to portal interactions of the mediator coupling the SM sector to the dark sector. The HV α_{dark} coupling is an important aspect of the model, to be denoted by α_{HV}. It can have a running value [147], with a starting scale.

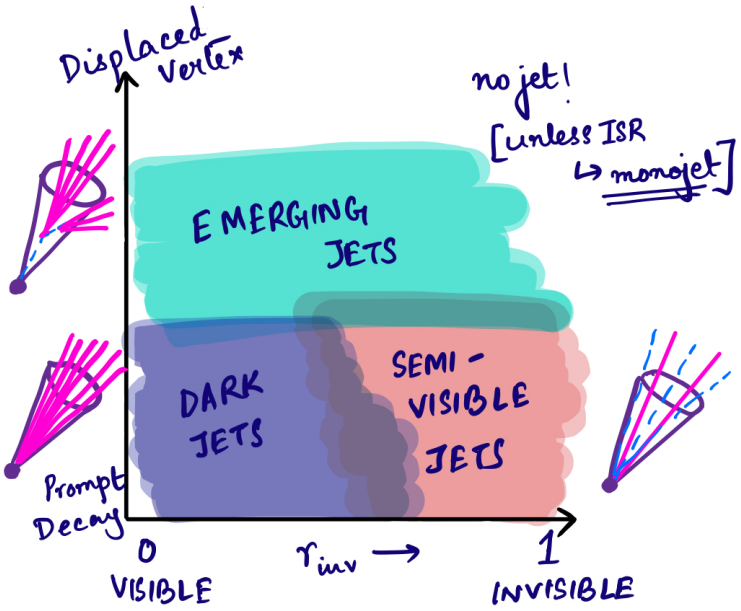


Figure 2.4: Schematic diagram showing evolution of jets with respect to different r_{inv} scenarios. Inspired by schematics drawn by Caterina Doglioni and Kevin Pedro.

Parameter	Value	Definition
HiddenValley:Ngauge	2	specifies the dark gauge group
HiddenValley:FSR	on	switch for having FSR shower on/off in a HV production process
HiddenValley:spinFv	0	specifies the spin of the HV partners of the SM fermions
HiddenValley:fragment	on	switch for having string fragmentation of the HV partonic system on/off
HiddenValley:pTminFSR	1.1	lowest allowed p_{T} of emission
HiddenValley:probVector	0.75	fraction of HV mesons that are assigned spin 1 (vector), with the remainder being assigned spin 0 (pseudoscalar)
HiddenValley:alphaOrder	1	Order at which α_{HV} runs
HiddenValley:alphaFSR	1.0	α_{HV} scale of g_v or γ_v emission

Table 2.2: PYTHIA8 HV parameter choices for the benchmark semi-visible jet model considered in this thesis. This list is in no way exhaustive and more details can be found in the PYTHIA8 HV webpage. All the parameters are kept fixed for event generation.

2.3.3 Semi-visible jets topology and extra jets

The uniqueness of t-channel mediator searches is due to the fact that there is no resonance to reconstruct, which is almost never the case for s-channel searches, and this leads to a more challenging final state. The event generation steps include the production of matched samples of $pp \to \bar{\chi}\chi +$ 0, 1, 2 jets respectively. The first process generated is that of pair production of DM particles ($pp \to \chi\chi$), the second process is that of associated production of DM particle and mediator ($pp \to \phi\bar{\chi}$), and the final process is that of pair production of the mediators ($pp \to \phi\phi^{\star}$, where $\phi = \chi j$).

The signal tends to have a large number of jets in most events, and that can be understood by looking at the event generation process. Fig. 2.5 contains an example of the Feynman diagrams produced during event generation of t-channel using MADGRAPH. There are contributions from cross diagrams even when generating t-channel solely, due to the presence of associated DM and mediator production modes and this increases the multiplicity of jets in the final state.

The necessity of having extra jets during the event generation arises from the fact that

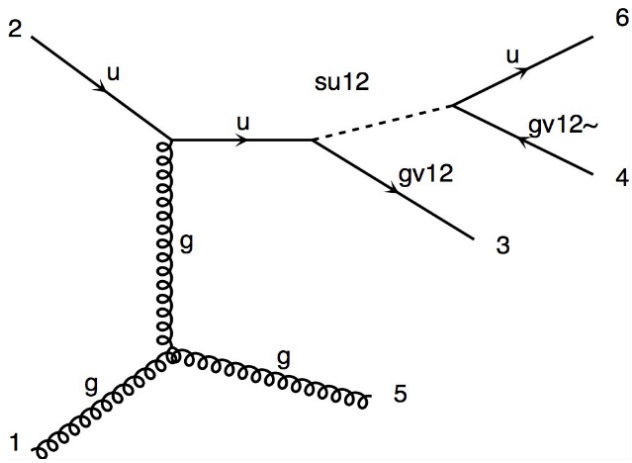


Figure 2.5: An example of a Feynman diagram showing cross contributions during t-channel production. The gv12 corresponds to the bi-fundamental mediator, whereas su12 is the dark quark in this case. u and g are SM quarks and gluons.

the DM pair production channel alone does not have sufficient amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ or hadronic activity and results in the production of back-to-back dark hadrons. The associated production of dark hadrons with a heavier mediator leads to the production of a second DM state, along with a hard parton (jet), which assures a larger amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ and hadronic activity, and the pair production of two mediators contribute even more to increase the hadronic activity significantly, thereby making the signal detectable over the Standard Model background.

Chapter 3

Collider basics and getting to know the ATLAS detector

The previous chapters discussed in detail the theoretical premise of this thesis, however, we are yet to discuss how the theoretical idea is translated into observable quantities, and what apparatus is required to achieve that feat experimentally. This chapter will focus on the basics of collider physics and delve into details about the ATLAS detector and associated collider aspects.

3.1 Large Hadron Collider at CERN

The Large Hadron Collider (LHC) is currently the most powerful particle accelerator, located at Centre for European Nuclear Research (CERN). The LHC is a circular proton-proton (pp) synchrotron collider, having a circumference of 27 km and a centre of mass energy of 13.6 TeV, with the start of Run-3 as of 5th July 2021. Apart from producing pp collisions, it also produces pA (A = lead) and AA collisions, however those collisions will not be further considered in this thesis.

The LHC provides collisions for four major and independent experiments, namely, ATLAS, ALICE, CMS and LHCb, spread along the circumference of the ring as shown in Fig 3.1. The layout of the ring is compartmentalised into octants, numbered in clockwise order, with each of the octants playing a significant role:

1. ATLAS experiment - multipurpose detector for pp, pA and AA collisions
2. ALICE experiment - multipurpose detector for pA and AA collisions, primarily focussing on heavy ion physics
3. Beam cleaning site
4. Radio Frequency (RF) cavities for acceleration of LHC beams
5. CMS experiment - multipurpose detector for pp, pA and AA collisions
6. Beam dumping site
7. Beam cleaning site
8. LHCb experiment - multipurpose detector for pp collisions, primarily focussing on heavy flavour physics.

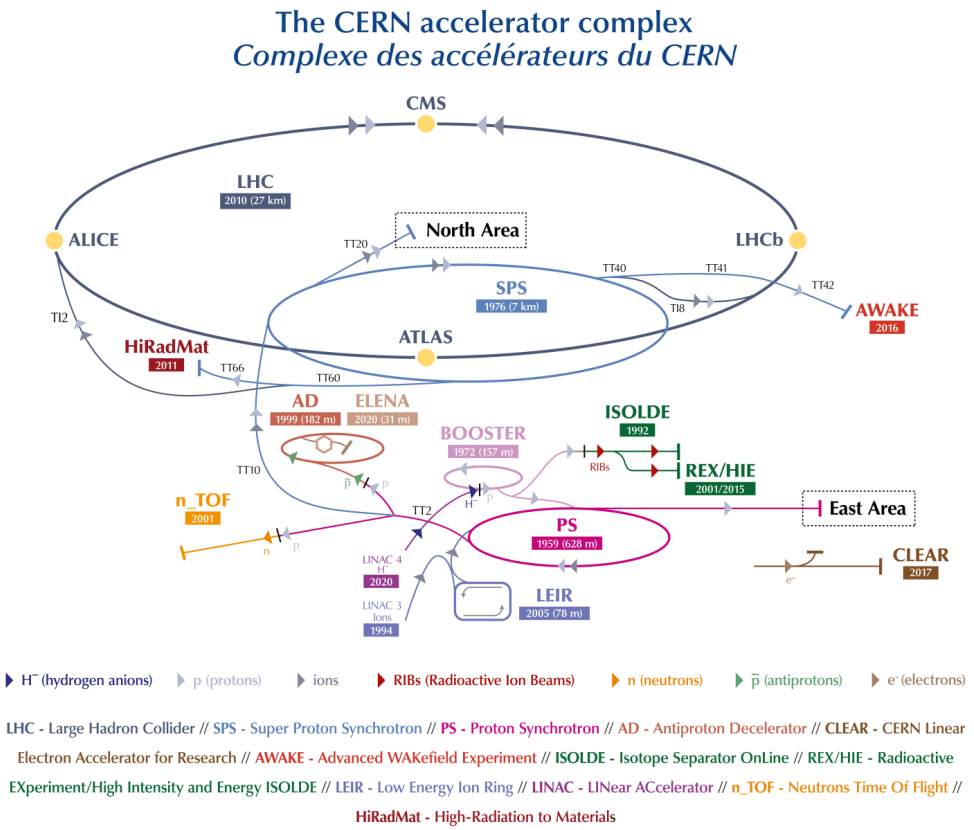


Figure 3.1: A schematic diagram of the CERN accelerator complex, which shows the multi-stage system of injectors which are used to fill the LHC. A subset of the many experiments supported by these accelerators is also shown [151].

The LHC is the final stage of a multi-step accelerator system, and as the starting point, an electric field is applied to hydrogen gas contained in a metal cylinder, which disintegrates the atom and creates protons to be fed into the subsequent steps. These protons are then accelerated in stages using a series of interconnected linear and circular accelerators – radio frequency quadrupole (QRF), linear accelerator (LINAC), proton synchroton booster (PSB), proton synchroton (PS), super proton synchroton (SPS) – ultimately entering the LHC ring. Within the LHC, two separate beams of protons are circulated in clockwise and counter-clockwise directions, that are set on a collision course by superconducting dipole magnets having a temperature of 1.9 K and 8.3 T magnetic field, which squeezes the beams making them narrower at the four interaction points along the ring [152]. The beams consist of *bunches* of protons, and the LHC is designed to run at 2808 bunches per beam, with the time difference between each bunch being 25 ns, creating ∼600 million collisions per second, and each collision is treated as an independent event.

The size and frequency of the proton bunches determine the rate of collisions. Due to the speed of the protons in the LHC (or any other hadron collider for that matter), the protons get Lorentz contracted longitudinally and can be thought to be oval-shaped. The instantaneous luminosity $\mathcal{L}$ is defined using the size of the bunches n_1 and n_2, the frequency with which bunches pass through each other f, and the beam cross-sectional area of overlap $\mathcal{A}$ as follows:

$$\mathcal{L} = \frac{f n_1 n_2}{\mathcal{A}} \tag{3.1}$$

$\mathcal{A}$ is related to the beam emittance ϵ and the beam cross-sectional size at the interaction β^* as $\mathcal{A} = 4\epsilon\beta^*$. The rate of occurrence of each individual physical process, is defined as the cross-section of that process. However, quantum mechanics becomes the "antagonist in the motion picture" and restricts us from accurately determining the outcome of each collision event, so in reality we derive the average rate of occurrence of a certain process, and label that as the cross-section of that particular process.

The number of events in a given amount of time and the cross-section of a desired process can be derived from $\mathcal{L}$ as follows,

$$\frac{dN_{events}}{dt} = \mathcal{L}\sigma. \tag{3.2}$$

The integrated luminosity is the the instantaneous luminosity integrated over all time, i.e. $\int \mathcal{L}dt$. The actual number of collisions for any particular process is given by,

$$N_{process} = \sigma_{process} \int \mathcal{L}dt \tag{3.3}$$

In order to maximize the sensitivity to new physics signals which typically have small cross-section, it is important to increase the integrated luminosity as much as possible, and that's where Eqn. 3.1 comes into play. Tweaking any of the parameters of that equation can increase the luminosity within the machine limitations, however increasing the luminosity using any of these techniques can lead to *pileup*. Pileup refers to the effect of extra pp collisions in the same or neighbouring bunch interactions, which contaminates the individual collision of interest. In-time pileup is parametrized in terms of the number of reconstructed primary vertices, N_{PV}, whereas out-of-time pileup is expressed in terms of the average number of additional pp colllisions per bunch crossing, μ. Fig 3.2 shows the progression of data collection, along with the total integrated $\mathcal{L}$ at the end of Run 2. It can be seen from Fig. 3.3 that the pileup increases with increased luminosity.

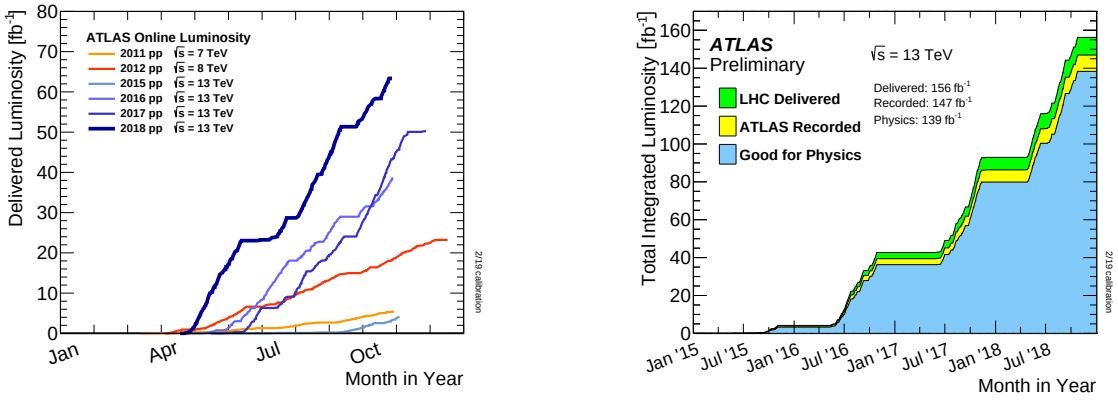


Figure 3.2: The cumulative luminosity versus day delivered to ATLAS during stable beams and for high energy pp collisions is shown in the left figure. The cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015-2018 is shown in the right figure [153].

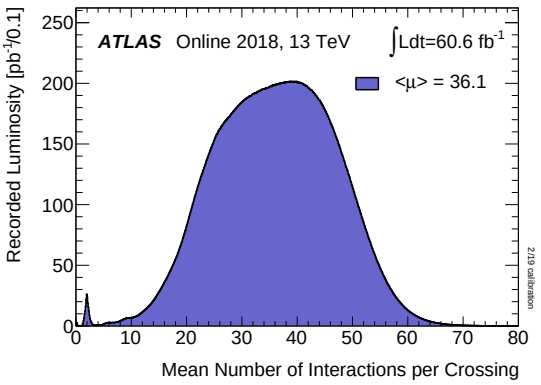


Figure 3.3: The luminosity-weighted distribution of the mean number of interactions per crossing for the 2018 pp collision data at 13 TeV centre-of-mass energy is shown. All data recorded by ATLAS during stable beams is shown, and the integrated luminosity and the mean mu value are given in the figure. It can be seen that the pileup increases with increased luminosity [153].

3.2 Give me the coordinates

In hadron colliders, although the momenta of incoming hadrons are known, the details of the interactions of the constituents (in terms of their energy and momentum fractions) are unknown. The collisions are characterised using the particles measured in the detector, and hence it is necessary to modify the standard spherical coordinate system to account for the cylindrical symmetry of the detector.

The incident proton beam direction is defined using the positive z-axis, and the origin of the coordinate system coincides with the collision point of the two beams at the centre of the detector. Only components along the x and y-axis can be measured, and the component of momentum in the transverse plane, $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$ is used. Considering the azimuthal angle, ϕ along the beam axis, $p_x = p_{\mathrm{T}} \cos\phi$ and $p_y = p_{\mathrm{T}} \sin\phi$.

A quantity termed as *rapidity*, y, is introduced instead of the conventional polar angle θ. It is defined in terms of the Lorentz-invariant particle kinematics, i.e. energy (E) and z-component of momentum (p_z) as,

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{3.4}$$

This can also be understood as $\cosh y = \gamma$, $\sinh y = \beta\gamma$, thereby implying that $\tanh y = \beta$. Here $\beta = v/c$ with v being the velocity of the particle, and $\gamma = \sqrt{1 - \beta^2}$ is the boost factor. They are the usual dimensionless quantities in relativistic kinematics since particle participating in or emerging from collisions within colliders move at speeds close to the speed of light. The detector geometry being cylindrical, each of the detector elements cover the same area in the $y - \phi$ plane, and it is easier for the rapidity to be mapped to that geometry.

Any boost in the system is parametrised by γ and by defining an axis. The corresponding change in energy and momentum along that axis is given by,

$$E' = E\cosh\gamma - p_z \sinh\gamma \qquad p'_z = p_z \cosh\gamma - E\sinh\gamma \tag{3.5}$$

Hence, $y' = y - \gamma$ by simple substitution of Eqn. 3.5 into Eqn. 3.4. This becomes important in cases where there exists a quark/gluon with a significantly larger value of momentum compared to the others in that collision. Inevitably all the particles will be

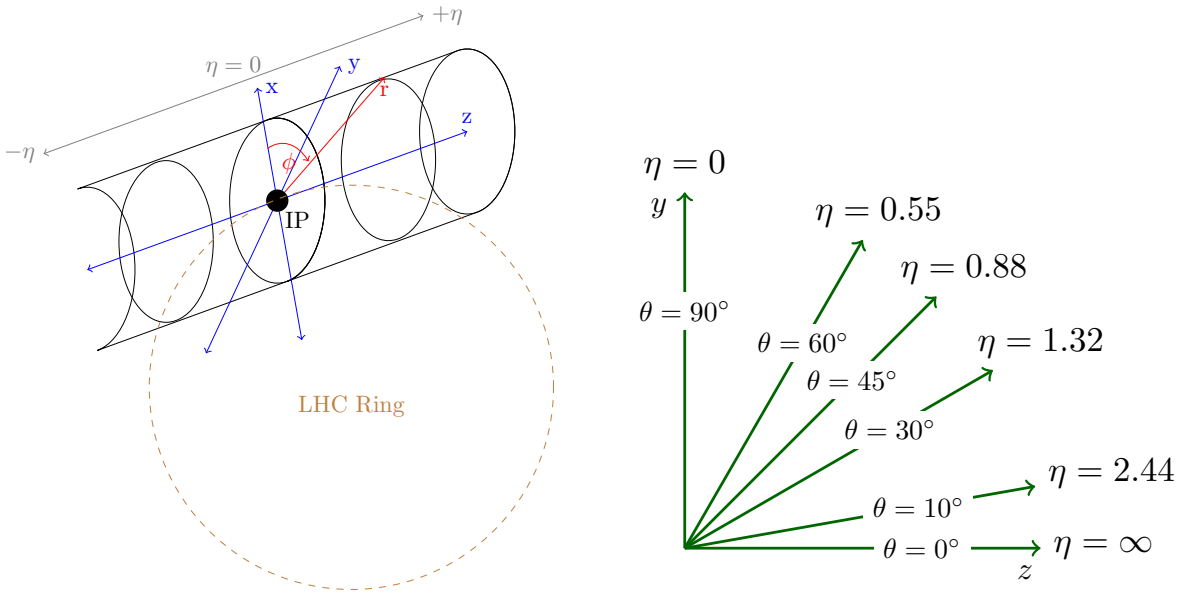


Figure 3.4: Diagram showing the coordinate system used in the LHC [27].

produced near one end of the detector, however, by looking at centre-of-mass frame rather than detector frame, the particles can be symmetrically distributed.

The z-component of momentum is not usually measured for detectors, since that it the beam axis, and hence rapidity can be difficult to measure for relativistic particles. For that purpose, another quantity, termed *pseudorapidity*, η, is devised, and it is defined as $\eta = -\ln \tan \theta/2$, using only the polar angle. η gives an indication of how far away from the center of the detector a particle gets produced, and can have values from $-\infty$ to $+\infty$ within the detector geometry. Fig 3.4 shows a schematic diagram of the coordinate system along with the conversion from θ to η. For massless particles, $y = \eta$, however, in case of massive particles, the two quantities do not coincide, and y is used in such cases.

The angle of emission of a particle is often described in terms of (y, ϕ) or (η, ϕ), and hence the angular separation between two particles can be expressed as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, which is equivalent to the standard opening angle in the spherical coordinate system.

3.3 Collisions – why do we love them?

Protons consist of quarks and gluons which interact with each other. When protons collide, this collision can happen either elastically or inelastically. In the former, new physics is not expected since the protons do not break up and there is no creation of new particles or loss of energy associated with such a collision, so nothing to "detect". In the case of inelastic collisions, where either or both of the protons can have a change in energy and direction, and are broken up, can typically result in the creation of new particles.

As shown in Fig 3.5, inelastic collisions can primarily occur in the following ways:

- Single diffractive - one of the protons dissociates into partons without having any direct interaction with the other proton.

- Double diffractive - both the protons dissociate into partons without any interactions between them.

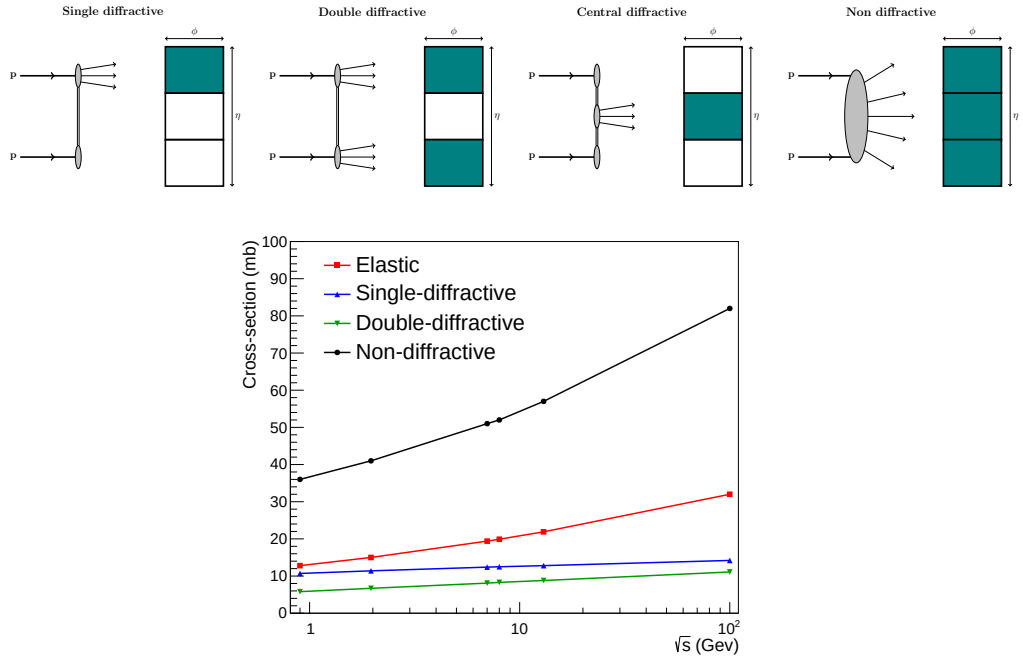


Figure 3.5: Schematic illustration of different types of inelastic collisions (left to right). The shaded areas in the rectangular representation of the detector in $\eta - \phi$ plane indicate if outgoing particles are observed there (top). The cross-sections corresponding to each inelastic process is shown with respect to centre of mass energy (bottom) [27].

- Central diffractive - it is assumed that diffraction is caused due to exchange of a colour-singlet hypothetical object called the Pomeron [154], and the proton dissociation occurs because hadrons get excited to a high mass state by absorbing the Pomeron and then decay. Both the colliding hadrons remain intact as they each emit a Pomeron, resulting in this process.

- Non diffractive - partons from two protons interact directly, effectively creating new particles. These collisions tend to have the most amount of energy exchange, and the outgoing particles are usually spread over the whole volume of the detector.

The total cross-section of events is a sum of elastic and inelastic collisions, with the inelastic component comprising of all the different types mentioned above.

3.4 Hello ATLAS detector

The ATLAS detector [155] is a general purpose hermetic detector located at LHC point 1. The motivation for the hermetic construction is to be able to cover all 4π of solid angle around the collision point. This enables reconstruction and identification of almost every energetic particle produced in an event, and the experiment is sensitive to pp, pA and AA collisions. The primary physics goals of the ATLAS detector are to discover and investigate the properties of the Higgs boson, conduct high precision SM measurements and search for new physics phenomena. The detector is capable of handling the expected high luminosity and pileup conditions, following several upgrades since its start in 2009.

The shape of the detector is cylindrical with several sub-detector layers and each layer wrapped around the previous one as shown in Fig 3.6. They can be broadly categorised as,

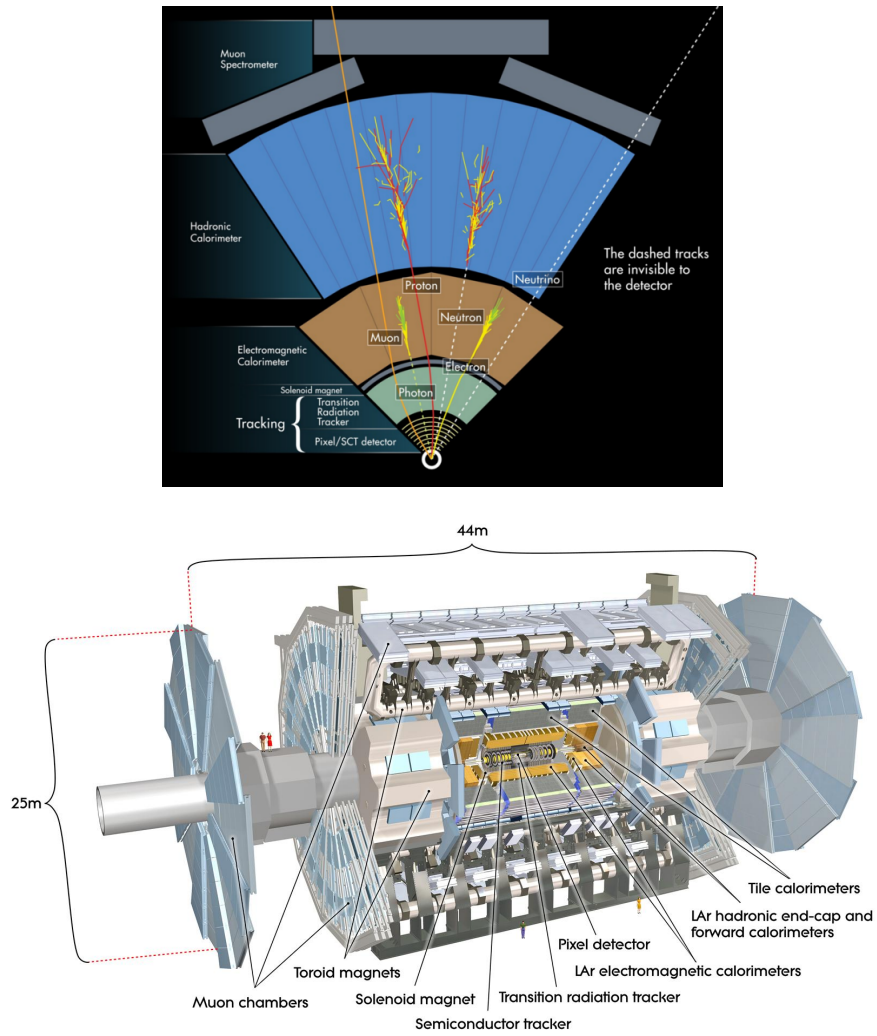


Figure 3.6: Schematic diagram of a generic detector, displaying different components (top), and a computer generated image of the ATLAS detector (bottom) [156]

- The tracking system: measures the charged particles emerging from the interaction point;
- The electromagnetic calorimeter: records the energy deposits of electro-magnetic particles like electrons and photons;
- The hadronic calorimeter: records the energy deposits of charged and neutral hadrons (pions and kaons) coming from quarks and gluons'
- The muon system: these tracking detectors record the interaction of muons since they are minimally ionising and hence do not interact with the EMCal material.

3.4.1 Inner detector

The innermost layers (comprising of three sub-sub-detectors) make up the tracking system covering up to $|\eta| < 2.5$, immersed within a magnetic field of 2 T from a solenoid, extending over a length of 5.3 m and 2.5 m diameter, as shown in Fig 3.7. In the barrel region, they are arranged as concentric cylinders around the beam axis, while in the end-cap regions they are located on disks perpendicular to the beam axis. Typically, tracks with p_{T} greater than

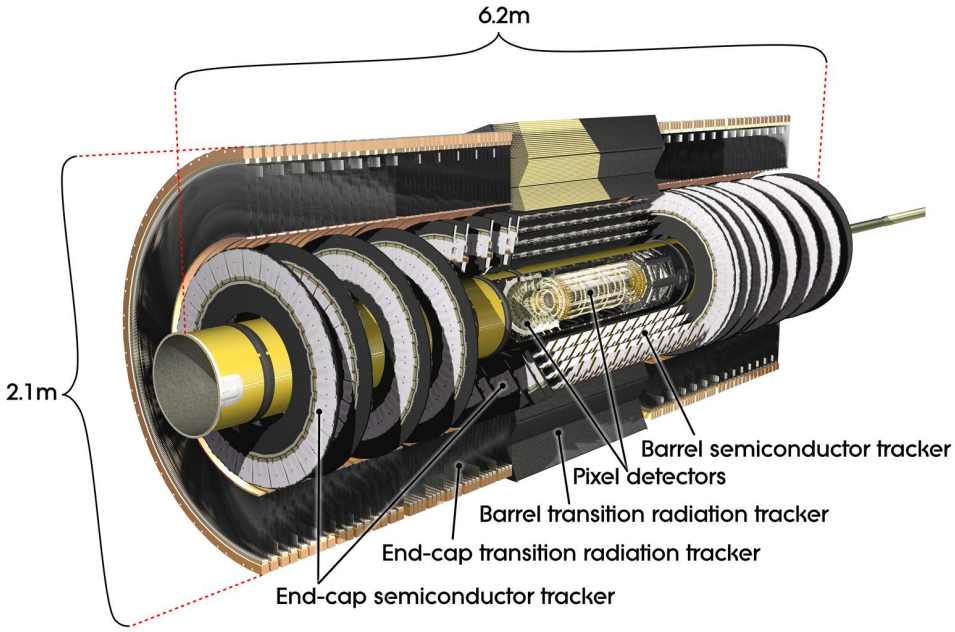


Figure 3.7: A computer-generated schematic representation of the ATLAS inner detector, with the various sub-detectors indicated [161].

0.5 GeV are used for most ATLAS analyses, however ATLAS can measure tracks down to a p_{T} of 0.1 GeV. The ATLAS inner detector system can be divided into three components:

- Pixel detector - consists of three cylindrical layers of high-granularity and high-precision semiconductor modules. Insertable B-Layer (IBL) [157, 158] is the inner most pixel layer in the ATLAS experiment, which was installed at 3.3 cm radius from the beam axis in 2014 to improve the tracking performance, and covers up to $|\eta| < 2.5$.
- SemiConductor Tracker (SCT) - contains layers of microstrip silicon wafers, arranged in four layers within the barrel and nine layers in the end cap. The barrel modules are available upto $|\eta| \approx 1.5$ and then the endcap modules take over [159].
- Transition Radiation Tracker (TRT) - a combined transition radiation detector and tracking detector, made up of layers of polyimide straw tubes both in the barrel and the endcap [160].

The tracking detectors are designed in such a way that the charged (and neutral) particles interact with least amount of material that can affect its path. Due to the presence of a magnetic field, charged particles have curved paths, with the sign of curvature denoting the charge and the momentum being estimated from the amount of curvature itself. Higher the p_{T} of the charged particle, less the curvature. Charged particles pass through the different layers of the tracking detectors and their position in each layer is recorded with high precision, termed as a *hit*. Pattern recognition algorithms are used to trace the hits and form a trajectory extrapolating back to the collision vertex, termed as *tracks*. The collision vertex is termed as *primary vertex* (PV) if it coincides with the assumed IP, and particles originating in PV are termed as prompt particles. On the other hand, particles originating from the decay of longer lived particles can be extrapolated back to vertices that do not coincide with PV and such vertices are termed as secondary vertices.

The tracks can be characterised using the transverse impact parameter, d_0 and the longitudinal impact parameter, z_0, both of which are measured with respect to the PV. d_0 is

the transverse distance from the PV position to the point of closest approach of the track in the $(\eta - \phi)$ plane, whereas z_0 is the distance along the z-axis. The momentum resolution of the tracking detectors, $\sigma_{p_{\mathrm{T}}}$, is defined as,

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} \approx \frac{p_{\mathrm{T}}}{BL^2} \times \sigma_s \tag{3.6}$$

where, σ_s is the uncertainty on the measurement of sagitta that gives the deviation of the track from a straight line, B is the magnetic field and L is the length of the track. As can be seen, the resolution can be optimised by either increasing the magnetic field, or the length of the tracking system. The most common charged particles are electrons, muons and pions. Muons being heavier have less curved and longer tracks compared to electrons. The fraction of a charged hadron being a pion is much higher than other baryons and mesons, and pions can leave spiral tracks in the tracking system.

3.4.2 Electromagnetic Calorimeter

The next layer after the tracking system is the electromagnetic (EM) calorimeter, as shown in Fig 3.8. Calorimeters are usually position sensitive, i.e. segmented, in order to measure the position of the energy deposit and the direction of the incoming particle. The EM calorimeter is designed to "stop" and record accurate energy measurements of any particle undergoing EM interactions, with the exception of neutrinos and muons (a minimally ionising particle at the LHC energy scale), that pass through the calorimeters mostly without depositing their energy. Particle(s) coming from the hard collision interact with the calorimeter material and create a cascade of particles, termed as a *shower*. EM showers are usually initiated by electrons/positrons or photons. At high energies, the electrons can emit photons via Bremsstrahlung radiation, whereas the photons can emit electron-positron pairs as well. At medium energies, the photons undergo Compton scattering which reduces their energy, and photoelectric effect converts these photons into low energy electrons, which are absorbed by the material to form ions, thereby ending the shower process. An electron is identified using the reconstructed tracks from the tracking detector, combined with localised deposits or clusters of energy in the EM calorimeter. Photons on the other hand, are identified using the localised energy clusters in EM calorimeter, along with the absence of any tracks in the tracking system leading to the clusters.

ATLAS calorimeters are sampling calorimeters and hence made up of alternating layers of active and passive materials. The passive material forces an interaction, leading to a shower, and the active layer accurately measures the amount of energy in the shower. The fraction of observed energy, i.e. the sampling fraction is defined as, $f_{sampling} = E_{active}/(E_{active}+E_{passive})$, and the ATLAS experiment has $f_{sampling} = 18\%$. The ATLAS EM calorimeter is made using Liquid Argon (LAr) as the active material, because of its radiation-hardness, stability and uniform nature. Lead has a short radiation length, and is chosen as the passive material, thereby enhancing the level of EM shower containment within the available space. Particles passing through the calorimeter ionises the LAr, and high voltages are applied to the plates enclosing the LAr, causing the electrons to drift to the copper electrodes.

The sub-detectors making up the EM calorimeter utilise an accordion like geometry, which reduces the number of blind regions of the calorimeter and provides full ϕ coverage.

- EM Barrel (EMB) calorimeter - covers $|\eta| < 1.475$, and contains three layers of different granularities and depths. The first layer has fine segmentation in η dimensions, and is useful in discriminating photons from $\pi^0 \to \gamma\gamma$ decays, as well as improving angular measurements. the second layer has fine segmentation in both η and ϕ space

and captures most of the energy in the calorimeter. The last layer collects the final remaining parts of the EM shower.

- EM EndCap (EMEC) calorimeter - covers the range of $1.375 < |\eta| < 3.2$ and contains two coaxial wheels on each side of EMB, aligned in the radial direction with a similar structure as EMB.

- EM presamplers - the $1.375 < |\eta| < 1.52$ range is the transition (*crack*) region between EMB and EMEC, and contains a significant amount of material that can aid to energy loss and reduced performance. To improve the energy measurements in this region, finely segmented LAr presamplers are placed in front of other material covering $|\eta| < 1.7$, which helps in reducing the impact of the crack region.

The quantity that describes the interaction between the detector material and the EM shower is termed as radiation length, X_0, which states the average distance of material needed to be traversed by an electron to reduce its energy to $1/e$ of the starting value. ATLAS EM calorimeter works with more than 22-38 X_0 of material to fully contain the EM showers.

The performance of a calorimeter can be estimated using the energy resolution, σ_E, which denotes the ability of a calorimeter to accurately determine the energy of an incoming particle, and it is usually defined as,

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \tag{3.7}$$

Here, E is the energy of the incoming particle, and a is a constant that depends on the calorimeter design. It is usually called the stochastic term since it is dominated by Poisson fluctuations, associated to the measurement of the shower. For the EM calorimeter, the full resolution can be obtained by adding a noise term and a constant term,

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E} \tag{3.8}$$

where, the notation $\oplus$ denotes addition in quadrature. The noise term accounts for quality of calibration, readout dead-time, non-linearity of response, whereas the constant term accounts for non-uniformity owing to pile-up contamination and other instrumental effects.

The ATLAS EM calorimeter [162] has a energy resolution of

$$\frac{\sigma_E}{E} = \frac{10\ \%}{\sqrt{E(\mathrm{GeV})}} \oplus 0.7\ \%$$

3.4.3 Hadronic Calorimeter

Hadronic showers tend to be more complex and longer compared to EM showers, and hence a separate calorimetry system is required to measure them. Roughly $1/3^{\mathrm{rd}}$ of hadronic interactions can produce π^0 meson, which can subsequently decay to two photons, thereby including an EM contribution to hadronic showers as well. The remaining $2/3^{\mathrm{rd}}$ of hadronic interactions can lead to a cascade of further hadronic activity, like the production of $\pi^{\pm}$ mesons, or the fragmentation of nuclei to protons or neutrons. There is a significant fraction of hadronic shower energy that goes undetected, primarily due to the loss of nuclear binding energy when an atom gets fragmented through strong interactions. An important

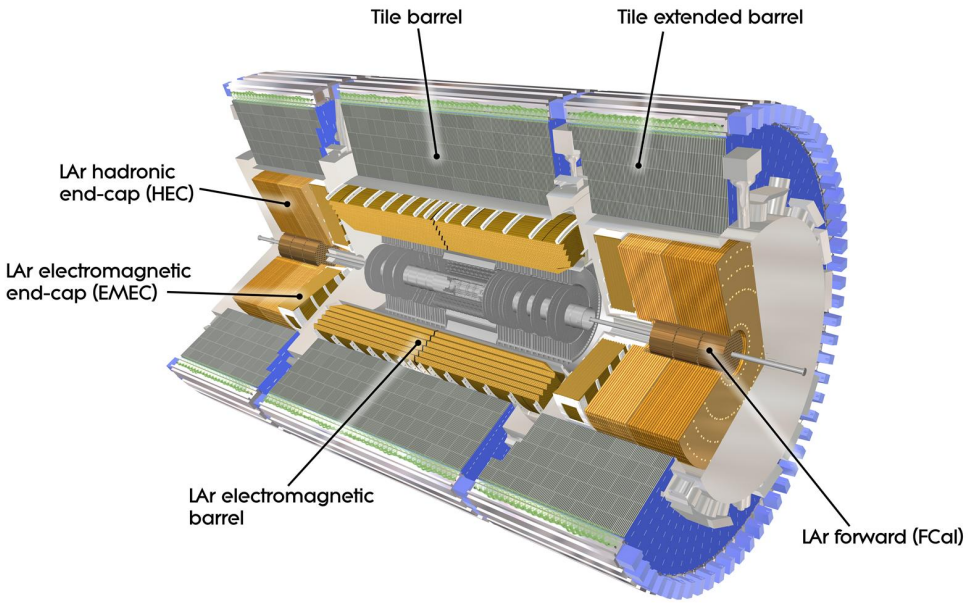


Figure 3.8: A computer-generated schematic representation of the ATLAS calorimeter, with the various sub-detectors indicated [163].

measure of how much a hadronic calorimeter can account for both EM and hadronic showers is the relative response of e/π ratio, where the numerator and denominator denotes the energy deposited by an electron and a pion respectively. If the $e/\pi = 1$, the calorimeter compensates for the additional shower particles. However, the ATLAS hadronic calorimeter is non-compensating, and hence cannot correct for this effect. This results in a response ratio of $e/\pi > 1$, which in turn implies that the energy response becomes non-linear with respect to energy, and it is corrected for, with a calibration procedure.

Hadronic calorimeters have to be larger than EM calorimeters. This is primarily due to the fact that for hadronic calorimeters the longitudinal development of the cascade is determined by the average nuclear interaction length, λ_I, which is the average distance that can be travelled by a hadron before it undergoes inelastic collision with the surrounding material. Atoms contain significant empty spaces and hence the nuclear interaction length can be larger than the radiation length. ATLAS hadronic calorimeter sub-components are designed to sustain depths of $7 < \lambda_I < 16$ cm, as shown in Fig 3.8.

- Tile Barrel and Tile Extended Barrel calorimeter - these segments use plastic scintillating tiles as the active medium, with the former covering $|\eta| < 0.9$ and the latter covering $0.8 < |\eta| < 1.7$. The absorbing layers have $\lambda_I = 16.8$ cm, and the tiles are arranged radially perpendicular to the beam line.

- Hadronic endcap calorimeter - similar to EMEC, however the passive absorption material in this case is copper. This component is comprised of two wheels with a coverage of $1.5 < |\eta| < 3.2$.

The energy resolution of the ATLAS hadronic calorimeter [164] is

$$\frac{\sigma_E}{E} = \frac{50\ \%}{\sqrt{E(\text{GeV})}} \oplus 3\ \%$$

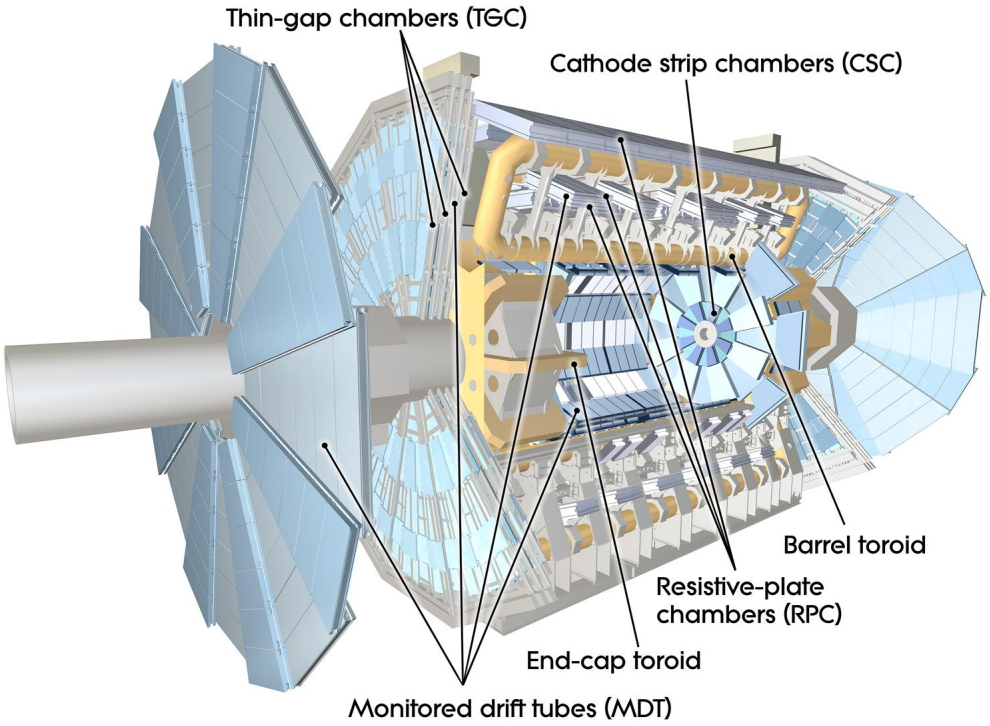


Figure 3.9: A computer-generated schematic representation of the ATLAS muon spectrometer, with the various sub-detectors indicated [165].

Forward calorimeter ATLAS forward calorimeter (FCal) is placed 4.7 m from the IP, in close contact with the end cap calorimeters, and has three sub components in itself – one EM calorimeter, and two hadronic calorimeters. The FCal covers a range of $3.1 < |\eta| < 4.9$. The active medium for all the three layers is LAr, however copper is used as the passive medium for the first layer, whereas the other two layers use tungsten. The main purpose of FCal is to record highly energetic particles, which do not deposit sufficient energy in the EMCal and/or HCal.

The energy resolution of the ATLAS forward calorimeter is

$$\frac{\sigma_E}{E} = \frac{100\ \%}{\sqrt{E(\text{GeV})}} \oplus 10\ \%$$

3.4.4 Muon spectrometer

In order to pass through the whole calorimeter system without losing its energy, a particle has to be either non-interacting, or minimally ionising. The former case is observed for the neutrino, and no detector component can tackle such a particle. However, the latter case is seen to be true for muons, and hence a dedicated set of detectors, collectively termed as the muon spectrometer (MS), is used for measuring this particle. Muons often provide very clean signatures since they are primarily reconstructed from their interactions in the MS alone.

The ATLAS MS, as shown in Fig 3.9, is immersed in a strong magnetic field generated using superconducting toroids in order to measure particle momenta using the curvature. It contains four sub-detectors, two in the barrel region (receiving 2.5 T magnetic field) and the other two in the endcap region (receiving 3.5 T magnetic field). One of the sub-detectors in each of the regions provide a coarse measurement or particle momentum, whereas the other

is designed to accurately determine the same with high precision.

- The coarse detectors cover $|\eta| < 2.4$ and are composed of Resistive Plate Chambers (RPCs) in the barrel region, and Thin Gap Chambers (TGCs) in the endcap region. These combined together form the fast muon system, due to their small time resolutions and ability to distinguish between different bunch crossings. The collection time is 15-25 ns making this system perfect for triggering purposes, discussed in more detail in the next section.

- The Monitored Drift Tubes (MDTs) are slower but precise detectors which cover $|\eta| < 2.7$ and are filled with a mixture of argon and CO_2. MDTs are installed in three layers and provide full coverage across the whole MS, however the MDT charge collection time is 700 ns, making it unsuitable to be used as a triggering system. Hence, additionally Cathode Strip Chambers (CSCs) are used for the innermost module of the innermost layer in the endcap region with $2.0 < |\eta| < 2.7$. CSCs have a maximum collection time of 40 ns, making it sensitive to the increased particle flux in this region.

3.5 Object reconstruction

Defining the topology is key to studying any physics process, and that requires defining the presence or absence of objects, which can either be a particle (like electrons, muons, taus, photons), a contribution from invisible processes leading to momentum imbalance in an event ($E_{\mathrm{T}}^{\mathrm{miss}}$), or a construction to understand the instrinsic hadron structure (jets, b-jets). All these objects are reconstructed from the detector tracks, cells and clusters, which can then be calibrated and applied to different analyses. The semi-visible jet analysis discussed in the next chapter primarily deals with jets and $E_{\mathrm{T}}^{\mathrm{miss}}$, and hence their reconstruction will be discussed in the following subsections.

3.5.1 Jets – how do we make them?

Jets are not fundamental objects, unlike the other physics objects mentioned above, but rather a construction to represent the hadronic shower originating from a single parton. Jets are essentially constructed when the collimated spray of particles are collected in a "single" object, by using different algorithms. Jets are defined by their inputs, algorithm and the radius parameter.

Jet inputs

Before diving into jet algorithms, it is important to understand what can be used as inputs to the algorithm. While there are several different input options, the one of relevance to this thesis is the Particle Flow (PFlow) approach. Traditionally ATLAS has relied solely on the calorimeter to reconstruct hadronic activity. PFlow was introduced as an alternate approach, where measurements from both the tracker and calorimeter simultaneously are combined and used as an input to jet algorithms. This was done because the former offers better angular resolution whereas the latter is good for energy resolution. The calorimeter energy deposits by charged particles are removed, and the PFlow objects are then essentially the remaining calorimeter energy and tracks that can be matched to the hard interaction, as can be seen from Fig. 3.10.

Topo-clustering is a process by which individual calorimeter cells are combined topologically to identify different collections of hadronic activity, as shown in Fig 3.11, and the

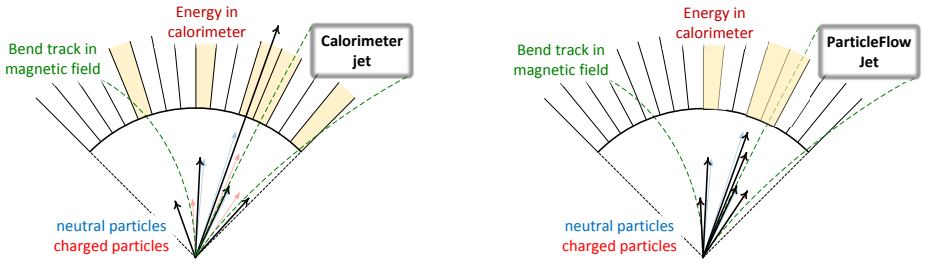


Figure 3.10: An illustration of the basic jet features for the calorimeter only configuration (left) and particle flow using tracks and calorimeter (right). The solid black arrows indicate the jet composition from representations of neutral (pale blue) and charged (pale red) particles by calorimeter towers or reconstructed tracks. The green dashed curves show charged particle tracks bend in the magnetic field. Both illustrations show the same truth-level jet [166].

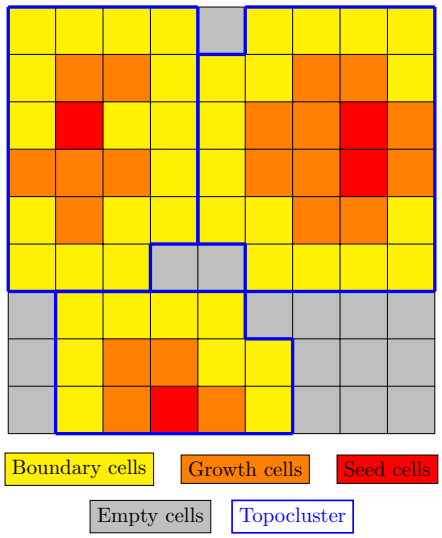


Figure 3.11: The energy deposits in the calorimeter cells can be combined in spatial dimensions to form clusters (right). The red seed cells, having the most energy compared to the neighbouring cells are chosen to initiate the process. The nearby orange growth cells, which have energy over a certain predetermined noise threshold are combined. Their neighbouring yellow boundary cells are added as well, to form the cluster indicated by the blue line. The grey cells have either no energy deposited in them, or have energy less than the noise threshold. Each cluster is formed dynamically depending on the position and energy of the cells.

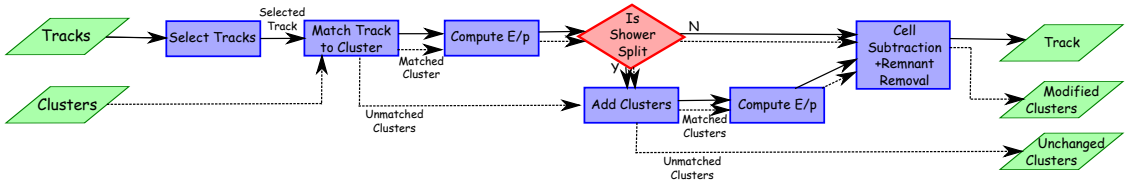


Figure 3.12: A flow chart of how the particle flow algorithm proceeds, starting with track selection and continuing until the energy associated with the selected tracks has been removed from the calorimeter. At the end, charged particles, topo-clusters which have not been modified by the algorithm, and remnants of topo-clusters which have had part of their energy removed remain, and are used as jet inputs [167].

resultant three-dimensional groups of cells are termed as topo-clusters. Along with inner detector (ID) tracks, these topo-clusters are the basic inputs to the PFlow algorithm. A cell-based subtraction scheme is used to remove overlaps between the momentum and energy measurements from the ID and calorimeters. The PFlow algorithm contains a list of tracks and topo-clusters (original and modified using the aforementioned energy subtraction scheme). A schematic of the PFlow algorithm is shown in Fig 3.12. Initially, well-measured tracks are selected with atleast nine hits in the ID, which the algorithm then attempts to match each track to an individual topo-cluster. The topo-cluster position and momentum are utilised to estimate the energy deposited in the calorimeter by the particle that contributed to the track. However, it is possible for a single particle to deposit its energy into multiple topo-clusters, hence that has to be taken into account by the algorithm. Depending on whether there is a probability of the shower being split or not, the algorithm decides if it has to add more topo-clusters to the track-topo-cluster system, in order to recover the complete shower energy. Then the energy associated with the selected tracks are removed cell-by-cell from the set of matched topo-clusters. Finally, a collection of charged particles, unmodified topo-clusters and modified topo-clusters remain.

It is advantageous to use PFlow objects as inputs to jet algorithms since they lead to improved energy and angular resolution of jets compared to the ones constructed only using the calorimeter information. The algorithm is found to be robust against pileup, outperforms a purely track-based pileup discrimination which was the ATLAS standard for several years, and consequentially performs better for hadronic observables. PFlow has been adopted as the standard ATLAS recommendation for Run 2, and an in-depth discussion of the PFlow approach implementation and performance can be found at Ref [167].

Jet algorithms

A typical jet algorithm can be considered as a mapping of a set of hadrons with four-momenta $\{p_1^{\text{had}}, p_2^{\text{had}}, ..., p_N^{\text{had}}\}$ to a set of jets with four momenta $\left\{p_1^{\text{jet}}, p_2^{\text{jet}}, ..., p_M^{\text{jet}}\right\}$, having $M < N$ and $p^{\text{jet}} = \sum_{i \in \text{jet}} p_i^{\text{had}}$. The jet axis in this case is defined by the resultant vector sum of the input four-momenta.

The jet algorithms should simple enough to implement in any experimental analysis while being complete enough to be consistent in a theoretical calculation. The algorithm should be definable at any order of perturbation theory, and be able to yield finite cross-sections which are relatively insensitive to the hadronization [168]. There are two broad classes of jet algorithms, namely,

- Cone algorithms [169] - a cone of fixed radius R is constructed around the input object with highest energy (corresponding to the jet axis), termed as the *seed* and all other

input objects within this cone is considered to be a part of the jet. The jet axis is then recalculated using all the input objects combined. This combination then becomes the new input seed, and the process is repeated until the point when there is no more change in the jet axis after recalculation. Then the next highest energy input is considered and a new iteration begins, and continues until all energy inputs have been used up. The major drawback of this class of algorithms is the possibility of having overlapping jet cones, being seeded by different inputs, as well as the lack of IR safety, as discussed previously in Chapter 1. To combat that issue, an alternative SISCone (Seedless IR Safe Cone) [170] approach has been developed, which finds distinct cones by exploiting the geometry of the detector.

- Sequential recombination algorithms - in these algorithms, the inputs are combined in momentum space instead of coordinate space as discussed above. Assuming that a jet is a repeated $1 \rightarrow 2$ splitting of partons, the jet can be obtained by repeatedly combining pairs of inputs. The original distance measure between the two inputs at each stage was defined as,

$$d_{ij} = \frac{2E_iE_j(1-\cos\theta_{ij})}{Q^2}$$

where, E_i and E_j are the energies of the two inputs and Q is the centre of mass energy. θ_{ij}) is the angle between inputs i and j. This measure was effectively the invariant mass of the two particle system scaled by the event energy scale. However, it caused problems for cases where two soft particles emitted at large angles would get combined to form a single jet. Additionally, this approach did not work out of the box for hadron colliders. Particles that go down the beam line are not measured, and therefore a beam distance parameter, d_{iB}, has to be included in the algorithm to account for the distance between in the beam and input object i. Furthermore, no fixed scale Q^2 can be obtained for a hadron collider since partons take part in the collisions, so a dimensionless angular parameter R_0 is a reference measure for the jet radius. Hence the distance measure was modified as,

$$d_{ij} = \text{Min}\left\{p_{T_i}^{2\rho}, p_{T_j}^{2\rho}\right\}\frac{\Delta R_{ij}^2}{R_0^2}, \quad d_{iB} = p_{T_i}^{2\rho}. \tag{3.9}$$

Here R_{ij} is the angular distance between the two input objects, and ρ determines the ordering or merging.

Due to the advantage of sequential algorithms over cone algorithms, when it comes to hadron colliders, the remainder of the discussion will focus on the same. The sequential jet forming ultimately boils down to the following steps:

1. The smallest value among all possible d_{ij} and d_{iB} is obtained.
2. If $d_{ij} < d_{iB}$, the inputs are merged, i.e. their p_{T} are added, and the p_{T} weighted average position is treated as the merged input position.
3. If $d_{iB} < d_{ij}$, the ith input is treated as a jet, and not considered in the calculation anymore.
4. The iteration continues until all inputs are categorised into jets.

Usual ATLAS values of R_0 is 0.4-0.6 for *small* jets, and 0.8-1.2 for *large* jets. Furthermore, the value of ρ can be used to create different jet algorithms as shown in Fig 3.13, such as,

- ρ = -1 (anti-k_t algorithm [171]): Clustering starts with the highest p_{T} input object, and all other softer objects are absorbed within that jet, for a given R_0. This leads to cone-like jets with the circular circumference of πR_0^2.

- $\rho = 0$ (Cambridge-Aachen algorithm [172, 173]): Inputs are clustered based on angular separation alone, irrespective of their momenta.

- $\rho = 1$ (k_t algorithm [174, 175]): Clustering starts with the softest input objects, and the jet outline changes when higher p_{T} objects are added to the cluster, resulting in an irregular jet shape.

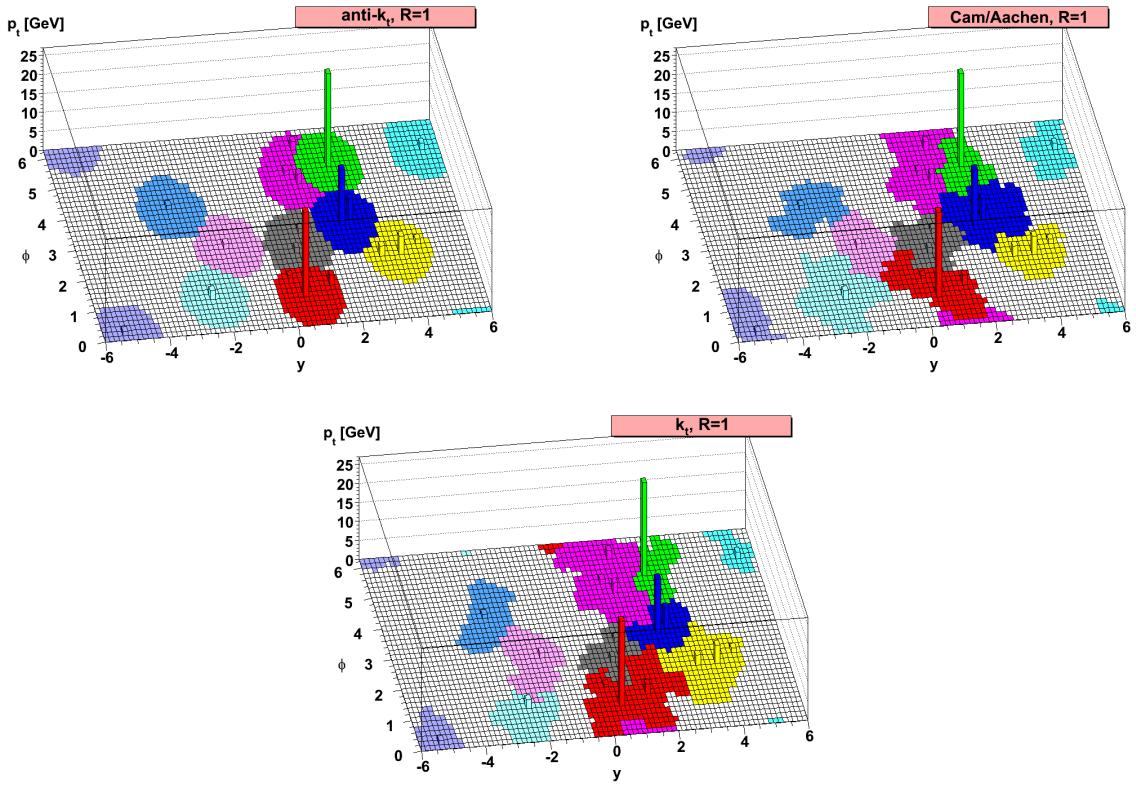


Figure 3.13: Jets obtained using anti-k_t (top left), k_{T} (top right), and Cambridge-Aachen (bottom) cone algorithms with R = 1. The colours have no physical meaning, they only served to distinguish energy deposits clustered together according to some algorithm [171].

The anti-k_t algorithm gives rise to circular jets, and hence is the preferred algorithm for the ATLAS experiment. The FASTJET [176] library provides an extensive collection of both the cone and sequential recombination family of jets, and is the default package used by the ATLAS jet reconstruction software.

The degree of collimation also plays an important factor when it comes to deciding the radius of a jet, due to the large amount of energy associated with the particles, termed as kinematic boost. As the centre of mass energy is increased, the jets become more energetic, resulting in more boosted systems. In physics processes involving hadronic decays of top quarks, W or Z bosons, it is useful if all of the decay products fall within a single jet. For example in $W \to qq$ or $Z \to qq$, both partons can be reconstructed as separate *small* jets (in terms of ΔR), however, assuming a *larger jet* with the two partons essentially being two *subjets*, can motivate walking back the chain and inferring the decay of the initiating

particle. Further, assuming that the energies of both the subjets are comparable, it can be concluded that they are in fact decay products of a W or Z boson. This approach is the baseline for the field of jet substructure, which will be discussed in detail in Chapter 5. The remainder of this thesis will however focus on small jets unless stated otherwise, as they are the primary objects in the semi-visible jet analysis.

3.5.2 $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction

Particles that interact neither via strong or electromagnetic interactions leave the detector without any visible signature. The energy-momentum conservation principle dictates that the total p_{T} of all particles after a collision should amount to zero, since the initial protons move along the beam axis. If there is an imbalance, it is attributed to the only SM particle that leaves no trace in the detector, i.e. neutrinos. This imbalance can arise in directions both parallel and perpendicular to the beam axis. However, the imbalance in the parallel direction is not consequential since the momentum fraction of each proton which participates in the collision is unknown, and most of the outgoing momentum parallel to the beam line remains unobserved. The primary quantity of interest is the momentum imbalance along the direction perpendicular to the beam line. It is reconstructed as the negative vector sum of of transverse momenta of all selected objects, as well as tracks compatible with the PV, and is termed as missing transverse momentum, with the magnitude being denoted by $E_{\mathrm{T}}^{\mathrm{miss}}$.

Due to its transverse nature, $E_{\mathrm{T}}^{\mathrm{miss}}$ can be built from its x and y components directly as,

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{\left(E_x^{\mathrm{miss}}\right)^2 + \left(E_y^{\mathrm{miss}}\right)^2}, \qquad \phi_{miss} = \arctan(E_y^{\mathrm{miss}}/E_x^{\mathrm{miss}}) \tag{3.10}$$

$E_{\mathrm{T}}^{\mathrm{miss}}$ is an event level quantity and requires the measurement of all physics objects. So, effectively, the $E_{\mathrm{T}}^{\mathrm{miss}}$ components can be calculated along the x and y axes as,

$$E_{x(y)}^{\mathrm{miss}} = E_{x(y)}^{\mathrm{miss},e} + E_{x(y)}^{\mathrm{miss},\gamma} + E_{x(y)}^{\mathrm{miss},\tau} + E_{x(y)}^{\mathrm{miss,jets}} + E_{x(y)}^{\mathrm{miss},\mu} + E_{x(y)}^{\mathrm{miss,soft}}. \tag{3.11}$$

The superscripts indicate the energy of the corresponding object. There can be extreme biases in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation if every quantity is taken at its face value, since electrons, photons and taus can also be mis-reconstructed as jets, which can lead to double counting, if these particles have already been identified before. To avoid double counting, overlap removal must be performed by matching energy deposits in the calorimeter and tracks to the reconstructed objects in the following order: electrons, photons, hadronically decaying τ, jets and muons. The components of the above equation associated with physics objects are referred to as the "hard" term, and those not associated with any physics object form the "soft" term.

Each of the hard terms have to be reconstructed independently, and have been listed briefly below.

- Electron term - electrons are reconstructed from clusters in the EM calorimeter which have an associated track in the ID, having a minimum $p_{\mathrm{T}} > 10$ GeV, and retained if they satisfy the "medium" selection criteria as listed in Ref [177].

- Photon term - photons are also reconstructed from clusters in the EM calorimeter which have an associated track in the ID, alongwith passing the "tight" selection criteria as listed in Ref [177]. To avoid double counting, in cases where the photon might already have been included as an electron, the electron that passes the above requirement is retained and the photon discarded.

- Hadronically decaying τ term - these objects are seeded by calorimeter jets in the central regions of the detector and have $p_{\mathrm{T}} > 20$ GeV.They are retained if they pass the "medium" selection criteria as listed in Ref [178].

- Jet term - anti-k_t jets with radius parameter R = 0.4 having $p_{\mathrm{T}} > 20$ GeVare used for this term. If it is seen that more than 50% of the jet energy is already considered by other overlapping objects, the remaining energy is accounted for by the soft term.

- Muon term - these particles are identified by matching ID track with MS tracks/segments, and are retained if they have a $p_{\mathrm{T}} > 5$ GeV.

The soft term is a particularly challenging component of $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. It is calculated using an *eflow* algorithm, which exploits the lower p_{T} threshold and good resolution of low p_{T} tracks to correct for low p_{T} topo-clusters as well as particles that could not reach the calorimeter. Although ATLAS utilises several approaches for soft term calculation, for the context of this thesis only the following two have been briefly mentioned. A more detailed explanation can be found in Ref [179].

- Calorimeter Soft Term (CST) - energy deposits in the calorimeter which are not matched to high-p_{T} objects make up this soft term. This algorithm also contains corrections based on tracks. Noise suppression is also implemented to reduce misreconstructed signals.

- Track Soft Term (TST) - tracks within ID that are not matched to high-p_{T} objects are utilised to make this soft term. This algorithm is very stable with respect to pileup, unlike CST, but does not include neutral particles.

The calorimeter and the tracking system in the ATLAS detector thus provide complementary information to the reconstruction of high-p_{T} objects as well as the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term.

3.6 Triggering on the objects

It is not possible to record and reconstruct every collision event due to limitations in data transmission and storage capacity, and hence it becomes necessary to have a trigger and data acquisition system that swiftly identifies events of interest and record only those. Deciding what can be categorised as an *interesting* event requires a certain level of care, so as to cover all possible topologies without biasing the trigger towards any specific process.

The harder the collision, the higher the probability of observing a high transverse momentum object, and hence triggers primarily target particles with high transverse momenta or events with low cross-section, since the cross-sections for interesting physics processes tend to be small. Currently, the LHC event rates are higher than 40 MHz, whereas the trigger rates are required to be below 1 kHZ, which corresponds to a factor of 10^4 reduction, and hence only a few hundred events are stored per second.

The ATLAS trigger system works in successive stages, termed as trigger levels. The guiding principle is that each stage of the trigger system stores a specific amount of events, which then becomes the input for the following stage. Thus, each successive stage has more time and information to optimise the decision. The different levels are shown in Fig 3.14 whereas an example trigger turn-on curve distribution is shown in Fig 3.15.

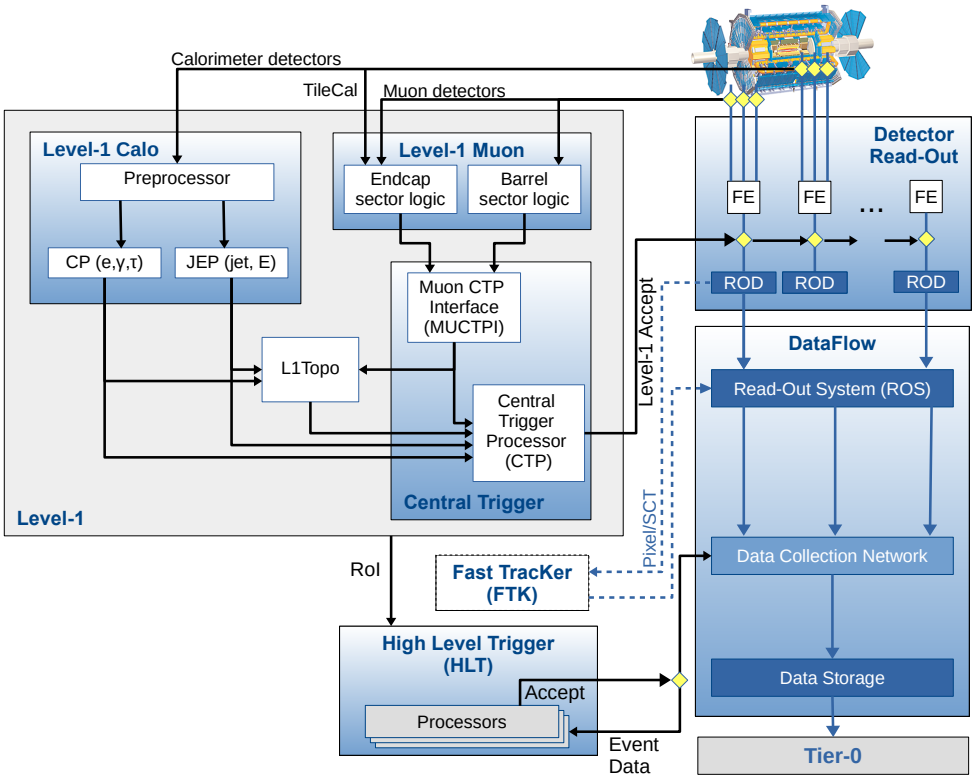


Figure 3.14: Schematic illustration of successive triggering steps [180].

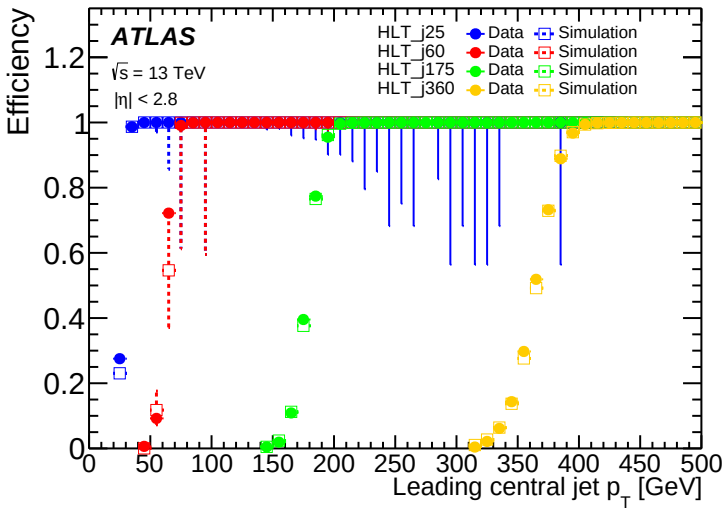


Figure 3.15: An example trigger turn-on curve plot [180].

Level 1 (L1) trigger operates with the tightest time constraints and demands some level of activity in the calorimeters and MS, since only these parts of the detector can perform simple reconstruction and trigger decisions on an event-by-event basis within 2.5 μs. The output is reduced from 40 MHz to ∼75 kHz at this stage. The L1 system uses calorimeter towers for reconstruction of electrons, photons and jets, whereas the muons are reconstructed using RPCs and TGCs with triggering capabilities. Depending on whether an event is deemed interesting by a L1 trigger, specific Regions of Interest (RoIs) are defined by the system in the (η, ϕ) plane, where the objects of interest are located. The RoIs also specify which trigger requirement was passed in order to "fire" the trigger.

High level trigger (HLT) receives the RoI information from L1 triggers, and can be used for regional reconstruction in trigger algorithms [180]. It has access to full detector granularity that lies within a RoI, and is a single homogeneous farm which allows for improved resource sharing and an simplified hardware and software dependencies. A full event reconstruction can be performed, using calibrations, and alignment corrections much like an *offline*[1] analysis, and reduces the event rate to the final level of approx 400 Hz. These events are then stored in a format that contains information of all of the digitised energy deposits from each of the detector components.

If the trigger is not efficient enough, it might throw out some events of interest. The trigger efficiency can be defined as,

$$\epsilon_{\text{trigger}} = \frac{N_{\text{accepted}}}{N_{\text{produced}}}.$$

Here the numerator gives the number of events containing the object of interest accepted by the trigger, and the denominator gives the total number of events with the object of interest. There are two ways of estimating the denominator. One approach would be to compare the number with respect to a looser trigger which is more inclusive in event selection. The other approach is to trigger on a different object, to estimate the selection efficiency on the object of interest. It is generally seen that there is a gradual rise of trigger efficiency, ultimately leading to the plateau at unity for any given trigger as shown in Fig 3.14, due to differences between object reconstruction at HLT level and offline level. These curves are termed as trigger *turn-on* curves.

Different processes have different probabilities of occurrence and that governs the frequency with which a trigger can be expected to observe that particular process. In order to address this problem, a *prescale* approach is employed, whereby after the trigger decides if a particular event is interesting, a random number is used to determine whether that event should be considered or not. This is decided using the prescale factor, $f_{prescale}$. For example, $f_{prescale} = 50$ implies that 1 out of every 50 interesting events is recorded. There can be scenarios where prescaling might not be beneficial, and in such cases, *unprescaled* triggers ($f_{prescale} = 1$) are used. This is the case for majority of ATLAS analyses since we are primarily interested in the high p_{T} regime.

The triggers should have a high efficiency for benchmark physics processes, since any event that is not triggered on gets lost forever[2]. With that idea in mind, a trigger menu is decided upon, which consists of a list of L1 and HLT triggers. There can be *primary* triggers that are usually unprescaled and used for physics analyses, however, *support* triggers are also

[1] If the object reconstruction is performed in real-time, i.e. during HLT trigger stage, it is termed as online. Any other reconstruction or analysis step that is carried out later on is termed as offline.

[2] A subset of this data is still stored with reduced information per event, and that is termed as delayed stream.

designed with prescale factors and are mainly used for efficiency and performance measurements or for monitoring. Apart from these basic ones, *calibration* triggers are designed and used for detector calibration, often being operated at very high rates but storing only the minimal information relevant for the calibration of the particular detector component. The trigger menu composition and thresholds are optimised for different luminosity ranges to maximise the physics output while remaining within the trigger bandwidth allowed by the ATLAS detector. Hence, sometimes dynamic prescales are used, which reduces the prescales as luminosity falls.

3.7 Detector simulation

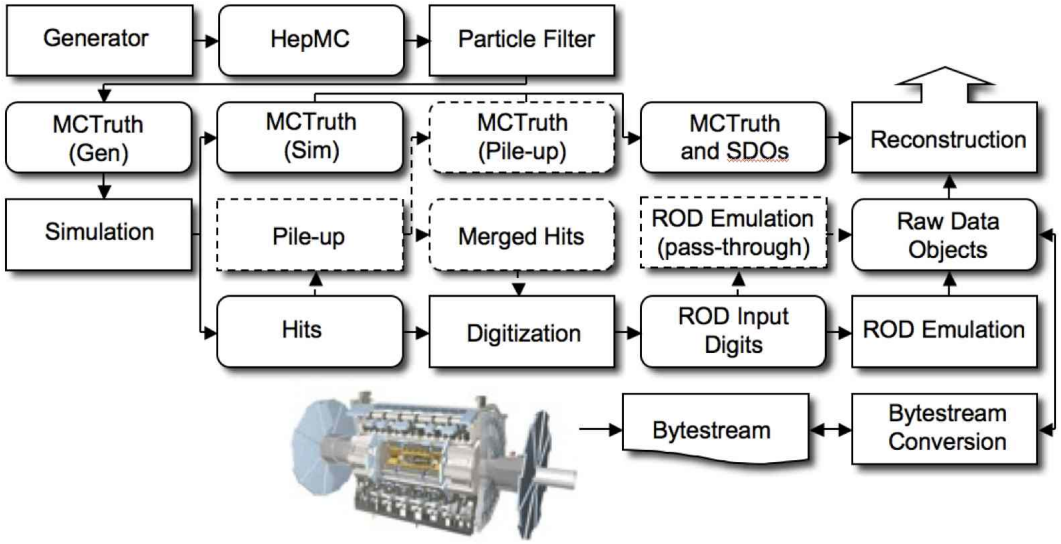


Figure 3.16: The flow of the ATLAS simulation software, from event generators (top left) through reconstruction (top right). Generators are used to produce data in HepMC format. Monte Carlo particle level information is saved in addition to energy depositions in the detector (hits). This particle level information is merged into Simulated Data Objects (SDOs) during the digitization. Also, during the digitization stage, Read Out Driver (ROD) electronics are simulated which are used to reconstruct physics objects [181].

Particle level events produced by Monte Carlo event generators have to be passed through detector simulation programs, in order to have the simulated events resembling the data obtained from collisions. The ATLAS simulation framework makes use of Geant4 toolkit for detector modelling [181]. There are several steps in the process, which are listed in Ref. [182]. Fig 3.16 shows the flow of the ATLAS simulation software, starting from event generation uptil reconstruction. The full simulation is computationally expensive, and hence an alternative fast simulation approach, termed as ATLFastII [181] is used when the full simulation precision is not necessary. During this process minimum bias events are overlaid on the hits to simulate the effect of pileup, taking into account the beam conditions, and physics objects are reconstructed.

3.8 Existing searches for Dark Matter with jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ in ATLAS

A broad overview of searches for DM at colliders is listed in Ref [183], and Fig. 3.17 shows the schematic of how the $E_{\mathrm{T}}^{\mathrm{miss}}$ is determined with respect to the visible particles in the collision for mono-X searches as discussed in Chapter 2, when viewed from the transverse plane.

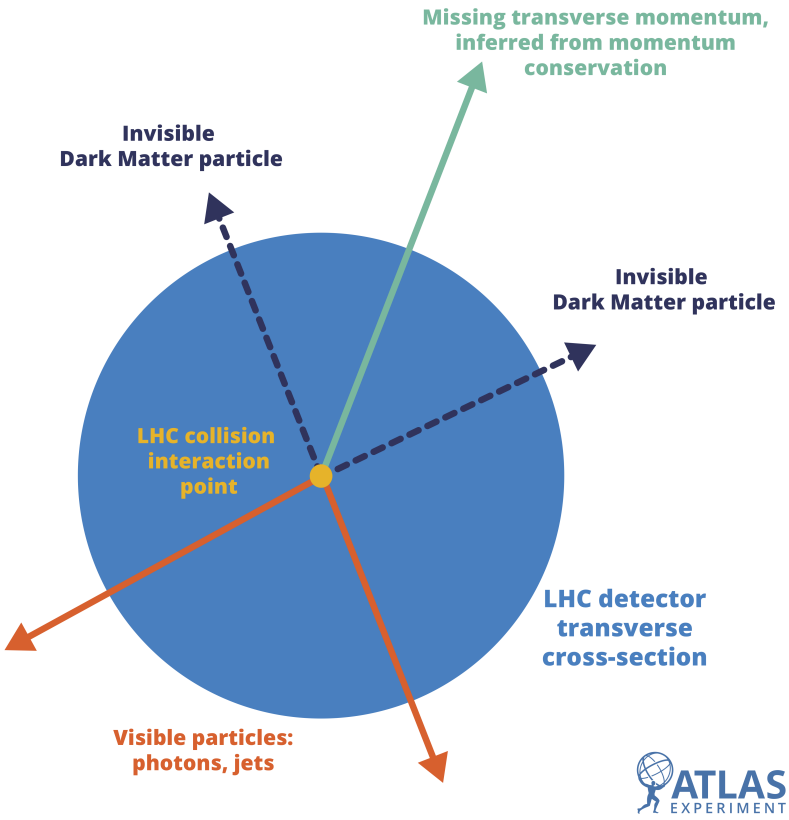


Figure 3.17: Schematic diagram showing how $E_{\mathrm{T}}^{\mathrm{miss}}$ is determined in the transverse cross-section of a LHC detector [184].

Mono-jet searches, as mentioned previously, look at events in the central region of the detector with a significant amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ and one or more energetic jets [185, 186]. The events are usually selected with at least one jet with p_{T} of 100-200 GeV and within the central region of the detector, along with a $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold of above 200 GeV. Restrictions are added on the hadronic and electromagnetic activity, to reduce the SM contributions and remove background contributions from non-collision events, in order to isolate signal-like scenarios. The dominant background contributions, following the different background reduction techniques, arise from $Z \to \nu\nu$ with jets and $W \to l\nu$ with jets. The number of events in signal enriched regions are looked at, for different $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds, and these signal regions can either be exclusive, i.e. $E_{\mathrm{T}}^{\mathrm{miss}}$ binned, or inclusive, i.e. taking into account all events after a certain $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold.

The advantage of such a search is that there are no specific assumptions about the nature of the DM particle(s), apart from the fact that they are produced in association with a SM jet. These searches can hence be used to look for a wide range of DM models, where a similar final state is expected. Since the SVJ final state involves jets and $E_{\mathrm{T}}^{\mathrm{miss}}$, it is natural

to test if the model can be partially or fully excluded using the limits set on the number of events by the existing ATLAS mono-jet search [185]. The signal samples as discussed in the next chapter were passed through the mono-jet event selection, and the number of events were compared to the existing limits. Fig. 3.18 shows the limits on the SVJ samples, discussed in detail in Chapter 4, as obtained by the mono-jet analysis, assuming a unity coupling strength between the SM quark, mediator and dark quark. As can be seen, as the mediator mass increases, and r_{inv} decreases, a substantial phase space cannot be excluded by the mono-jet search.

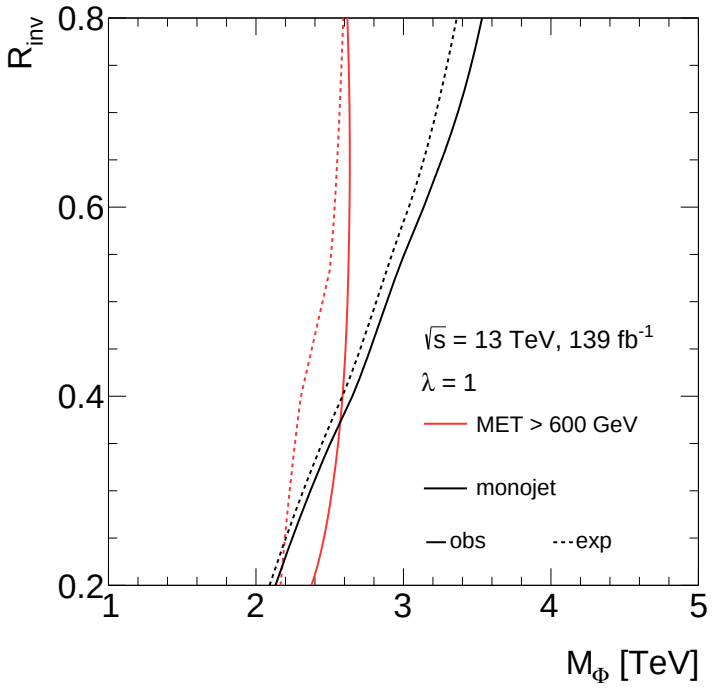


Figure 3.18: Limit on semi-visible jet signals obtained from mono-jet selection, for different mediator masses and r_{inv} values.

Searches with photon + $E_{\text{T}}^{\text{miss}}$ [187, 188] and vector boson + $E_{\text{T}}^{\text{miss}}$ channels [189–192] can be studied alongside the standard jet + $E_{\text{T}}^{\text{miss}}$ channel when the recoil arises from dark interactions instead of ISR, but they tend to have smaller and different backgrounds and uncertainties. Uptil now, jet + $E_{\text{T}}^{\text{miss}}$, photon + $E_{\text{T}}^{\text{miss}}$, vector boson + $E_{\text{T}}^{\text{miss}}$ or Higgs boson + $E_{\text{T}}^{\text{miss}}$ [193, 194] searches in ATLAS and CMS have been able to set constraints on several different simplified models of DM mediation, however many more theoretical scenarios and phase-space are yet left to be explored.

Chapter 4

Search for non-resonant production of semi-visible jets in ATLAS

This chapter is based on Ref [195], where the thesis author was the main analyzer and INTERNAL/CONF note editor. All ATLAS material (unless stated otherwise) is from this reference. The paper is in the final stages of preparation.

4.1 Context

Collider searches for Dark Matter (DM) until this date have mostly focussed on scenarios where DM particles are produced in association with either heavy Standard Model (SM) particles, or jets. However, no confirmed evidence of DM has been observed so far. Several models [136, 139, 147, 148] have been proposed that include a strongly-coupled dark sector as introduced in Sec. 2.3, giving rise to unusual and unexplored collider topologies. Semi-visible jets (SVJ) [143] is one such example, and a search for the t-channel production mode is presented in this chapter. Searches for the t-channel production mode probe a broad class of non-resonant signals and can potentially have higher mass reach, as they are not limited only to finding resonance peaks as in the s-channel searches.

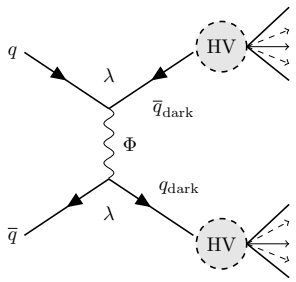


Figure 4.1: Illustrative Feynman diagram and subsequent production mechanism of semi-visible jets via a t-channel mediator, Φ, producing a pair of dark quarks, marked by q_{dark}. HV denotes the Pythia8 Hidden Valley module used to simulate interactions connecting the dark sector with the SM sector, which produces a final state consisting of SM hadrons and dark hadrons, governed by the r_{inv} fraction. The coupling strength of the q-q_{dark}-Φ interaction is denoted by λ.

The SVJ signal model (described in Chapter 2) predicts an event topology where the event $E_{\text{T}}^{\text{miss}}$ is aligned with one of the jets. In the t-channel production mode, a scalar bi-fundamental mediator (Φ) acts as a portal between the SM and dark sectors. It couples to

a SM quark and a dark quark and mediates the production of dark quarks in pp collisions, as shown in Fig. 4.1.

The ratio of the rate of stable dark hadrons over the total number of hadrons coming from the PYTHIA HV shower in the event is termed r_{inv}, which is a free parameter of the model. This results in reconstructed jets geometrically encompassing the dark hadrons, termed semi-visible jets. At leading order the two SVJs are back-to-back and the direction of the missing transverse momentum is aligned with one of the two reconstructed jets. However, this signature is dominated by background events from dijet processes. A boost by additional jets leads to signatures with the $E_{\text{T}}^{\text{miss}}$ usually pointing in the direction of one of the two SVJs, since both of them contribute to the $E_{\text{T}}^{\text{miss}}$. On the contrary, for QCD mutijet processes the $E_{\text{T}}^{\text{miss}}$ typically originates from one severely mis-measured jet and therefore the $E_{\text{T}}^{\text{miss}}$ is more aligned with the direction of one of the jets. As this is a signature of mis-measured jets in a detector, these events are typically discarded in searches involving jets and $E_{\text{T}}^{\text{miss}}$. Most searches use stringent selection requirements to suppress the multijet background, which is why these scenarios are mostly unexplored at the Large Hadron Collider (LHC).

The SM background contributions will be discussed in Sec. 4.2. This is followed by the object definitions discussed in Sec. 4.3. The dataset and simulation details are provided in Sec. 4.4. Sec. 4.5 highlights the event selections and region definitions, whereas the general analysis strategy is presented in Sec. 4.6. Sec. 4.7 and 4.8 discusses the associated systematic uncertainties and background estimation strategy employed in this analysis, whereas Sec. 4.9 describes the simultaneous fit strategy. Finally, Sec 5.3.2 and 5.5 discusses the results and conclusions.

4.2 Standard Model background processes

The SVJ signal has an unique final state, however, there are SM background processes that can lead to a similar final state either due to an object being mis-reconstructed or mis-identified. This section will present an overview of the dominant SM background processes contributing to this search, and the percentage quoted for each background corresponds to its contribution in the signal enriched region which will be discussed in Sec. 4.6.

The primary source of SM backgrounds that can result in semi-visible jet like events is the $Z \to \nu\nu$ with jets final state. This is an irreducible background because of the weakly interacting nature of neutrinos. The neutrinos provide a source of real $E_{\text{T}}^{\text{miss}}$, which makes this final state signal-llike. For this analysis, $Z \to \nu\nu$ + jets accounts for 50% of the total SM background contribution, and hence it is necessary to estimate this background as precisely as possible.

The second largest background contributions come from the $W \to l\nu$ with jets processes, where again the presence of a real neutrino contributes to true $E_{\text{T}}^{\text{miss}}$, making the final state signal-like. These processes are reducible backgrounds because of the presence of a real lepton, that aids in identifying and rejecting these events. When $l = e, \mu$, the event can be construed as signal-like if the lepton is mis-identified or outside of detector acceptance. However, if l is a τ, there can be two possibilities. If the τ decays leptonically, similar situation as for e, μ is expected. If the τ decays hadronically, the resulting jet can easily be interpreted as a signal jet. This family of processes is estimated to contribute roughly 35% of the total background.

The third dominant backgrounds are the $t\bar{t}$ and single-t productions, standing at roughly 10% of the total SM background contribution. There can again be two possibilities of these processes being interpreted as a signal-like event – having unidentified or non-reconstructed leptons in the final state when the top quark decays leptonically (including τ), or having jets

from the case where the top quark decays hadronically. In the former, there are contributions from true $E_{\mathrm{T}}^{\mathrm{miss}}$ due to the presence of neutrinos in the final state.

The multi-jet background can result in signal-like events due to the presence of energetic jets which can be mis-reconstructed or mis-measured, leading to fake $E_{\mathrm{T}}^{\mathrm{miss}}$. This process should in principle be the dominant background for the semi-visible jet final state, due to its large production cross-section. However, after specific kinematic selections which will be discussed in detail later on, multi-jet background can be significantly reduced and is seen to comprise roughly 5% of the total background.

There are sub-dominant contributions arising from diboson (WW, ZZ or WZ) + jets processes, especially in cases where either or both of the bosons can decay to two jets, or two leptons, or a lepton and a neutrino, the last option contributing to real $E_{\mathrm{T}}^{\mathrm{miss}}$. However, the production cross-section of these processes is sufficiently low at LHC energies and hence they comprise less than 4% to the total SM background for this analysis. $Z \to ll$ + jets can also contribute to this final state, however the threshold of the contribution is minimal because in this case both the leptons have to be either mis-reconstructed, or arise from hadronically decaying τ, or are out of detector acceptance.

Finally, the non-collision background (NCB) is expected to be a contributing source as well, because this background is associated with large amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ and energetic jets. Dedicated cleaning criteria have to be imposed in order to reduce this background to negligible levels in the different regions of the analysis, as will be discussed in detail later on.

4.3 Object definitions

The SVJ analysis uses jets, electrons, muons, τ-jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$ for either selecting or vetoing events. The reconstruction and selection criteria of the objects are listed below.

Jets jets are reconstructed using the anti-k_t [171] algorithm with radius parameter $R = 0.4$, which takes as input PFlow objects, that are charged particle tracks matched to the nominal primary vertex with the requirement $|z0 \sin\theta| < 2.0$ mm[1] and calorimeter clusters surviving an energy subtraction algorithm that removes the calorimeter deposits of good-quality tracks from any vertex [196]. These jets are calibrated with a series of steps that first correct for pileup contamination, and then correct the average reconstructed jet energy response to match that of jets formed from stable particles after hadronisation. Response corrections include an overall MC-based Jet Energy Scale (JES), a Global Sequential Calibration to correct for residual dependence of the jet response on the shower development and initiating parton flavour, and a final *in situ* correction applied only to data [197].

[1] z0 is the longitudinal distance of the PV with the closest point of approach of the track.

Feature	Criterion
Pseudorapidity range	$\|\eta\| < 2.8$
Transverse momentum	$p_{\mathrm{T}} > 30$ GeV
Algorithm	anti-k_t
R-parameter	0.4
Input constituent	EMPFlow

Table 4.1: Summary of jet reconstruction criteria.

Feature	Criterion
Jet selection	$p_{\mathrm{T}} > 30$ GeV
	$\lvert\eta\rvert < 2.8$
Algorithm	BDT DL1r
Operating point	Eff = 77%

Table 4.2: Summary of b-tagging selection criteria.

Feature	Criterion
Transverse momentum	$E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV
Algorithm	Track-based (TST)
Soft-term	Track-based (TST)

Table 4.3: Summary of $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction criteria.

Jets pass the baseline criteria if they satisfy $p_{\mathrm{T}} > 30$ GeV, $|\eta| < 2.8$. The Jet vertex fraction (JVF) is defined as the fraction of tracks in the jet associated with the PV, and jet vertex tagger (JVT) [198] uses JVF to obtain the tagged jets with a certain threshold. To suppress jets originating from pileup collisions, requirements on the JVT [199], discriminant are applied for jets with p_{T} below 60 GeV. Scale factors are applied to correct the (in)efficiencies of selecting simulated jets with JVT to those assessed in data. A summary of the definition is listed in Table 4.1. An alternative study was performed to check if reclustered jets [200, 201] can be a viable option for this topology. The results and the reason why it was not ultimately selected have been described in App. A.1.

B-tagged jets baseline anti-k_t $R = 0.4$ jets described above are tagged as b-jets if they pass the 77% efficiency working point of the *DL1r* algorithm [202]. A 77% efficiency b-tagging operating point corresponds to mis-tag rate of charm jets lying between 2% to 15% dependent on jet p_{T} and that of light-flavour jets lie at the level of 1% [203]. Events with two or more b-tagged jets are vetoed to reduce $t\bar{t}$ background contributions. A tighter veto cannot be applied as the signal contains some fraction of b-tagged jets. Scale factors are applied to correct the (in)efficiencies of selecting simulated b-jets to those assessed in data [204]. A summary of the definition is listed in Table 4.2.

Missing transverse momentum the $E_{\mathrm{T}}^{\mathrm{miss}}$ is rebuilt from the selected PFlow jet collection [205]. As the jet selection needs to be applied to fully calibrated jets, and the corresponding jet energy scale (JES) calibration is not applied during reconstruction, $E_{\mathrm{T}}^{\mathrm{miss}}$ needs to be rebuilt in the derived data, by applying the correct JES calibration and valid p_{T} threshold. After the desired set of hard object collections have been handled, the jet term and soft term(s) are simultaneously used to build $E_{\mathrm{T}}^{\mathrm{miss}}$ using a calorimeter or track based jet term, respectively. The *tight* WP is used in this analysis, where central jet $p_{\mathrm{T}} > 20$ GeV, forward jet $p_{\mathrm{T}} > 30$ GeV and JVT cut is 0.5 for jets up to 60 GeV. A summary of the definition is listed in Table 6.3.

Leptons electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that are associated with charged-particle tracks reconstructed in the inner detector. Electrons are required to fulfill 'tight likelihood' identification criteria as well as

calorimeter and track based isolation criteria [206]. Muon candidates are reconstructed by combining inner detector tracks with muon spectrometer tracks or energy deposits in the calorimeters consistent with the passage of muons. Muons are required to fulfill 'medium' identification criteria as well as calorimeter- and track-based isolation criteria [207]. Electrons and muons are required to satisfy $p_{\mathrm{T}} > 7$ GeV and be within the tracking volume $|\eta| < 2.5$. Electrons are also required not to be in the transition region between the barrel and endcap EM calorimeters, given by $1.37 < |\eta| < 1.52$.

Hadronically-decaying τ-lepton candidates are formed by combining information from the calorimeters and inner tracking detectors with a Recurrent Neural Network algorithm, and they are required to have one or three associated tracks [208]. They are required to pass *medium* identification requirements, to have $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 2.5$ excluding the transition region between the electromagnetic barrel and endcap calorimeters and to have one or three associated charged tracks. Events with reconstructed τ candidates are vetoed.

Primary vertex the vertex with the highest $\sum p_{T,track}^2$, where $p_{T,track}$ is the transverse momentum of tracks associated with the vertex, is chosen as the primary vertex. The events must have at least one vertex with at least two tracks.

Overlap removal the jet algorithm groups all topo-clusters without any further requirements of whether or not any given topo-cluster has already been used in the reconstruction of another type of object. Hence, it is expected that jets and other objects can share the same energy deposits, and this double-counting of energy is known as overlap. Overlap removal (OLR) must be performed for a consistent treatment of the event kinematic quantities and to avoid ambiguities in reconstructing the $E_{\mathrm{T}}^{\mathrm{miss}}$. An overlap removal procedure, following the usual sequence of steps [191] is applied in the SVJ analysis, and listed in Table 4.4. Initially, jets close to leptons are removed is they satisfy the selection criteria, but if there is a further jet close to a lepton after this omission, then the lepton is considered to be coming from semi-leptonic decays of heavy flavour jets and the lepton is removed.

Remove	Keep	Matching criteria
Muon	Electron	Muon with calorimeter deposits and shared inner detector track
Electron	Muon	Shared inner detector track
Jet	Electron	$\Delta R < 0.2$
Electron	Jet	$\Delta R < \min(0.4, 0.04 + 10\,\mathrm{GeV}/p_{\mathrm{T}}^{e})$
Jet	Muon	Number of tracks < 3 and $\Delta R < 0.2$
Muon	Jet	$\Delta R < \min(0.4, 0.04 + 10\,\mathrm{GeV}/p_{\mathrm{T}}^{\mu})$

Table 4.4: Overview of the overlap removal between objects and the corresponding matching criteria, listed according to priority.

4.4 Dataset and Simulation

4.4.1 Simulated samples

The signal samples are generated with MadGraph5_aMC@NLO [209] event generator at LO in QCD, using the *DMSimp* model from Ref. [210], with up to two extra partons at leading order, and showered with Pythia8 [211, 212] with the NNPDF2.3LO [213] parton distribution function (PDF) set and the A14 tune [214]. The samples were generated with r_{inv}

values of 0.2, 0.4, 0.6 and 0.8, with dark hadron mass of 10 GeV, and mediator mass of 1000-5000 GeV. The cross-sections for each of the mediator mass points are listed in Table 4.5. PYTHIA8 Hidden Valley [215] module specific parameters are as listed in Table 2.2. Standard matching/merging was applied in order to avoid over counting of jets in an event.

Mediator mass [GeV]	LO cross-section [fb]
1000	9470
1500	964
2000	131
2500	32
3000	14
3500	7
4000	4
4500	2.5
5000	1.6

Table 4.5: Details of signal samples generated with r_{inv} values of 0.2, 0.4, 0.6 and 0.8, with dark hadron mass of 10 GeV, each with full detector simulation.

The contributing background processes are W/Z+jets, $t\bar{t}$, single-t, multi-jet, and diboson, as listed in Tab. 4.6. The MC simulated samples are processed through a detailed ATLAS detector simulation [219] based on Geant4 [220], and then reconstructed and analysed using the same procedure and software that are used for the data. Additional simulated pp collisions generated using PYTHIA 8 with the A3 set of tuned parameters [221] and NNPDF2.3LO PDF set are overlaid to simulate the effects of additional collisions from pileup.

4.4.2 Data samples

The data sample used in this analysis, collected during normal operation of the detector, corresponds to an integrated luminosity of 139 fb^{-1} collected in 2015, 2016, 2017 and 2018. The 2015 dataset was reprocessed during 2016 using 2016 reconstruction software. In each year of data taking, a subset of the recorded data is generally unsuitable for using in an analysis due to detector problems or other imperfect data taking conditions. When such conditions occur, the associated events are tagged in an ATLAS database, and added to a

Table 4.6: Summary of generators used for simulation of background processes, along with the PDF and tune used.

Process	Generator	ME order	PDF	Parton shower	Tune
W/Z+jets	SHERPA2.2.11 [216, 217]	NLO (up to 2 jets)	NNPDF3.0NNLO [218]	SHERPA MEPSatNLO	Default
$t\bar{t}$	POWHEG BOX2 [45, 66, 218]	NLO	NNPDF3.0NLO	PYTHIA8.230 with NNPDF2.3LO	A14 [214]
Single top	POWHEG BOX2	NLO	NNPDF3.0NNLO	PYTHIA8.230 with NNPDF2.3LO	A14
Multijet	PYTHIA8.230 [212]	LO	NNPDF2.3LO	PYTHIA8.230	A14
Diboson	SHERPA2.2.1	NLO (up to 2 jets)	NNPDF3.0NNLO	SHERPA MEPSatNLO	Default

list which specifies that they should be excluded from typical analysis selections. Data is only used if it passes the final Good Run List (GRL) released by the ATLAS Data Quality group.

4.5 Event selection and cleaning

Once the objects have been defined, the next step in any analysis is deciding the event selection criteria. For a search, the event selection should be performed so as to enhance the signal sensitivity and simultaneously reducing background contributions. The SVJ analysis targets an unusual phase space of having $E_{\mathrm{T}}^{\mathrm{miss}}$ aligned along one of the jet directions while being as inclusive as possible, and hence the event selection is optimized to generically reject backgrounds and retain significant signal statistics in the high $E_{\mathrm{T}}^{\mathrm{miss}}$ regions.

As discussed previously, there can be situations where events are corrupted due to unsuitable detector conditions, i.e. components of ATLAS detector not being fully functional, or objects have been mis-reconstructed or mis-identified. These cases have to be taken into account when selecting events, and hence appropriate event cleaning measures have to be employed.

4.5.1 Triggering strategy

The triggering criteria decides which collision events are recorded, and hence is a crucial step for selecting events specific to an analysis. The SVJ analysis uses unprescaled triggers purely based on $E_{\mathrm{T}}^{\mathrm{miss}}$, since the signal is expected to have high $E_{\mathrm{T}}^{\mathrm{miss}}$ values. The $E_{\mathrm{T}}^{\mathrm{miss}}$ in the trigger is based only on calorimetric measurements and does not include any reconstructed muons, so the muons behave similarly to invisible particles in this trigger [222].

Different algorithms are employed at the HLT level to categorise and accept events. All algorithms account for the energy measured by the calorimeter and it is associated with some set of energy depositions, generally referred to as elements, however the definition of the set of elements is algorithm-dependent.

- The *cell* algorithm is the most basic HLT algorithm, which calculates $E_{\mathrm{T}}^{\mathrm{miss}}$ as the sum across all calorimeter cells with energy deposits with a pre-decided threshold considering the effects of pileup contributions and electronic noise.

- The *mht* algorithm calculates $E_{\mathrm{T}}^{\mathrm{miss}}$ from the negative transverse momentum vector sum of all jets with $p_{\mathrm{T}} > 7$ GeV. This algorithm is especially useful, since in most cases, jets tend to dominate the visible momentum, and are usually corrected for pileup effects.

- The *pufit* algorithm corrects for pileup effects based on the assumption that high $E_{\mathrm{T}}^{\mathrm{miss}}$ deposits are associated to a hard scatter whereas low $E_{\mathrm{T}}^{\mathrm{miss}}$ deposits originate from pileup.

In 2015, a threshold of 70 GeV was used, which was subsequently raised multiple times to cope with increasing effects from pileup, reaching 110 GeV during the 2017-2018 data-taking period for *mht*. The threshold for *pufit* algorithm was set at a constant value of 110 GeV, however an additional requirement on the *cell* $E_{\mathrm{T}}^{\mathrm{miss}}$ of 50 GeV and 65-70 GeV was applied for 2017 and 2018 data respectively.

Trigger efficiency

The lack of muon information in the HLT trigger definition is utilised in order to estimate the trigger efficiency. A detailed study is performed to verify that selected events are safely on the plateau of the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger turn-on in MC and data. The efficiencies of $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers are derived in a single-muon event sample obtained with the single muon triggers. The same selection as described in Section 4.5, except that all $E_{\mathrm{T}}^{\mathrm{miss}}$-related requirements are dropped. While the V+jets background consists mostly of Z+jets processes rather than W+jets, it is found that the trigger response for these two processes is nearly identical, so that an estimation of the trigger efficiency in W+jets events is also valid for Z+jets.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency is defined by:

$$\text{efficiency} = \frac{\text{number of events passing selection AND } E_{\mathrm{T}}^{\mathrm{miss}} \text{ trigger requirement}}{\text{number of events passing selection}} \tag{4.1}$$

The efficiencies are calculated for each $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger and are also evaluated separately for data and MC. The $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger is found to be fully efficient for $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV. As a result no trigger efficiency scale factors or uncertainties are applied.

4.5.2 Data quality

Events are not considered by the analysis if they fail one of the following criteria: LAr, Tile and SCT error requirements. These preselections are performed to remove detector errors, and are only applied to data samples. Incomplete data events are identified from the event core flag, and removed. In the selected events, there are no SCT, Tile and LAr calorimeter problems and also no lost detector fragments in the detector readout [185]. Additionally, in all data periods some dead Tile calorimeter modules are present. Jets pointing close to these regions have mostly underestimated energy measurement, which results in increased $E_{\mathrm{T}}^{\mathrm{miss}}$ value. The effect of dead Tile modules is only partially simulated in MC, and since for accurate SM background estimate for this analysis, it is essential to reduce the influence of the dead Tile regions via a set of fiducial cuts based on the angular orientation of the jet, following Ref. [223].

4.5.3 Jet Cleaning

Events with jets containing anomalous energy depositions due to coherent noise and electronic noise bursts in the calorimeter [224, 225] are removed, but this has a negligible effect on the signal efficiency. Non-collision backgrounds (NCB), for example energy depositions in the calorimeters due to muons of beam-induced or cosmic-ray origin, are suppressed by imposing an additional selection criteria on the leading jet: the ratio of the jet charged-particle fraction to the maximum fraction of the jet energy collected by a single calorimeter layer, $f_{\mathrm{ch}}/f_{\mathrm{max}}$, is required to be larger than 0.1. Events are rejected if they contain a selected jet pointing in the direction where tile calorimeter modules were disabled, or other detector errors were present, and the studies have been listed in App. A.2. The NCB contribution is also found to be negligible after these requirements.

4.5.4 Event selection

The SVJ analysis utilises one signal region (SR) and multiple control regions (CRs) and this section aims to motivate the necessity of each of the regions. The SR is designed to enhance

the presence of the SVJ signal, and the different CRs are selected in such a manner that negligible signal contribution is expected in those regions. The CRs can hence be used to estimate (i.e. control) the contribution of a given SM background.

The following selections are applied for the analysis, with the justifications presented subsequently. Tables 4.7 and 4.8 depicts events surviving at each stage of this preselection process for four benchmark signal samples and the dominant background samples. In general, higher r_{inv} values results in events with higher $E_{\text{T}}^{\text{miss}}$, leading to a higher signal acceptance.

For the analysis, the primary event preselection requirements are as follows:

- Events are required to pass the $E_{\text{T}}^{\text{miss}}$ triggers, and have $E_{\text{T}}^{\text{miss}} > 200$ GeV.
- Events with any electrons or muons or hadronic τ candidates, passing the respective selection criteria, are vetoed.
- Events are required to have at least two jets with R=0.4, with the leading jet $p_{\text{T}} > 250$ GeV, other jet $p_{\text{T}} > 30$ GeV and $|\eta| < 2.8$. The recommending et cleaning criteria is applied on all jets, with a stricter requirement on the leading jet, as discussed in Ref. [226].
- Events with two or more b-tagged jets are vetoed to suppress $t\bar{t}$ background.
- Events are required to have at least one jet within $\Delta\phi < 2.0$ of $E_{\text{T}}^{\text{miss}}$ direction. We will use $\Delta\phi(\text{closest jet}, E_{\text{T}}^{\text{miss}})$ to denote this angle.

Step/Process	Signal (1500)	Signal (2500)	Signal (3500)	Signal (4500)	All MC
XS × Lumi	83337.4	1628.82	169.859	40.7247	1.03317e+07
dead-tile correction	77962.9	1538.42	161.514	39.2275	9.94517e+06
$\Delta\phi(\text{closest jet}, E_{\text{T}}^{\text{miss}})< 2$	76430.4	1494.66	151.997	35.7988	8.36492e+06
$p_{\text{T}}^{\text{lead}} > 250$ GeV	76147.5	1477.3	146.296	33.6423	5.7206e+06
$n_{\text{b-jets}} \leq 1$	68215	1312.81	133.01	31.3642	4.91719e+06
$n_{\tau\text{-jets}} < 1$	67738.9	1304.75	132.034	31.1753	4.79896e+06

Table 4.7: Cutflow table for four benchmark signals of different mediator masses in GeV and $r_{\text{inv}} = 0.4$ and combined background contribution from MC samples. All MC implies the total background estimated from all simulated samples, including multi-jet.

Step/Process	2000 (0.2)	2000 (0.4)	2000 (0.6)	2000 (0.8)	All MC
XS × Lumi	6848.67	10169.9	11811	12116.1	1.03317e+07
dead-tile correction	6385.45	9548.97	11124.3	11552.5	9.94517e+06
$\Delta\phi(\text{closest jet}, E_{\text{T}}^{\text{miss}})< 2$	6335.19	9353.7	10593.2	9998.42	8.36492e+06
$p_{\text{T}}^{\text{lead}} > 250$ GeV	6308.32	9308.86	10537.4	9922.44	5.7206e+06
$n_{\text{b-jets}} \leq 1$	5324.42	8193.18	9659.62	9364.7	4.91719e+06
$n_{\tau\text{-jets}} < 1$	5295.48	8139.87	9607.64	9318.59	4.79896e+06

Table 4.8: Cutflow table for four benchmark signals of mediator mass 2000 GeV with different r_{inv} and combined background contribution from MC samples. All MC implies the total background estimated from all simulated samples, including multi-jet.

Three independent control regions are defined, to assess the background estimation methodology. This analysis use muons as they offer a cleaner handle on the relevant processes with less mis-reconstructed lepton contributions. The different region definitions are summarised in Tab. 4.9.

Event preselection for 1L region the single lepton (1L) selections are used to estimate the contributions from W+jets, and top quark processes. This is motivated by the fact that the SR contamination from these backgrounds is dominated by W-boson decay to τ leptons, comprising roughly 61% of the total W+jets contribution. The primary event selection requirements for 1L region are as follows:

- Events are required to pass the $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers, and have recalculated $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV, by adding the selected muon.
- Events with one muon with $p_{\mathrm{T}} > 7$ GeV are selected. Events with electrons and hadronic τ-leptons are vetoed.
- Events are required to have at least two jets with R=0.4, with leading jet $p_{\mathrm{T}} > 250$ GeV, other jet with $p_{\mathrm{T}} > 30$ GeV, and $|\eta| < 2.8$.
- Events are required to have at least one jet within $\Delta\phi(\text{closest jet}, E_{\mathrm{T}}^{\mathrm{miss}}) < 2.0$.
- Events with one or more b-tagged jets are vetoed to suppress $t\bar{t}$ background.

Event preselection for 1L1B region for a estimation of the background contribution from top processes (semi-leptonic $t\bar{t}$ and single top quark induced processes), an additional event selection requirement is imposed along with the 1L selection. This requires events to have one or more b-tagged (1B) jets, leading to 1L1B region. The primary event selection requirements are as follows:

- Events are required to pass the $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers, and have recalculated $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV, by adding the selected muon.
- Events with one muon with $p_{\mathrm{T}} > 7$ GeV are selected. Events with electrons and hadronic τ-leptons are vetoed.
- Events are required to have at least two jets with R=0.4, with leading jet $p_{\mathrm{T}} > 250$ GeV, other jet with $p_{\mathrm{T}} > 30$ GeV, and $|\eta| < 2.8$.
- Events are required to have at least one jet within $\Delta\phi(\text{closest jet}, E_{\mathrm{T}}^{\mathrm{miss}}) < 2.0$.
- Events are required to have at least one b-tagged jet.

Event preselection for 2L region finally, for the Z+jets processes estimation, a 2L selection is employed which selection ensures a pure $Z \to \mu\mu$ with jets sample. The two opposite sign muons are treated as neutrinos, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is recalculated accordingly. To eliminate contributions from top processes, an additional requirement is imposed, which demands the events to have no b-tagged jets. This is because $Z \to \nu\nu$ with jets and $Z \to \mu\mu$ with jets should have same kinematics, and the additional jets give events with $E_{\mathrm{T}}^{\mathrm{miss}}$ direction close to a jet. The primary event selection requirements are as follows:

- Events are required to pass the $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers. The two opposite sign muons are then considered as neutrinos, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is recalculated. The events are required to have recalculated $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV.
- Events with exactly two opposite sign muons of $p_{\mathrm{T}} > 7$ GeV, within the Z-boson mass window ($60 < M_Z < 120$ GeV) are selected. Events with electrons and hadronic τ-leptons are vetoed.

Region	Nominal (0L SR)	1L CR	1L1B CR	2L CR
Trigger	$E_{\mathrm{T}}^{\mathrm{miss}}$			
Lepton Requirement	Electrons and muons (7 GeV) == 0	Muon (7 GeV) == 1		Muon (7 GeV) == 2, opp charged
Recalculation of $E_{\mathrm{T}}^{\mathrm{miss}}$	None	Adding muons to $E_{\mathrm{T}}^{\mathrm{miss}}$		
b-tagged jets	< 2	== 0	>= 1	No requirement

Table 4.9: Summary of event pre-selections for different SR and CRs.

- Events are required to have at least two jets with R=0.4, with leading jet $p_{\mathrm{T}} > 250$ GeV, other jet with $p_{\mathrm{T}} > 30$ GeV, and $|\eta| < 2.8$.
- Events are required to have at least one jet within $\Delta\phi(\text{closest jet}, E_{\mathrm{T}}^{\mathrm{miss}}) < 2.0$.
- Events are required to have no b-tagged jets.

In all these selections using muons, the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger is still used to select the events, but the $E_{\mathrm{T}}^{\mathrm{miss}}$ is recalculated considering muons to be invisible, to mimic the $E_{\mathrm{T}}^{\mathrm{miss}}$ definition of the nominal analysis.

4.6 Analysis strategy

The main observables used in this analysis are listed below:

- jet multiplicity, n_{jets}
- missing transverse momentum, $E_{\mathrm{T}}^{\mathrm{miss}}$
- the scalar jet p_{T} sum, H_{T}
- the leading jet p_{T} , $p_{\mathrm{T}}^{\mathrm{lead}}$
- the angle between the closest jet (termed svj) and $E_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta\phi(\text{closest jet}, E_{\mathrm{T}}^{\mathrm{miss}})$

Following the event preselections, it is necessary to identify a phase space governed by different kinematic variables, that can help enhance the sensitivity of the signal, and also estimate the SM background. The signal events are expected to have at least two jets, with high $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions. One of the key features of this signal is the presence of a jet aligned along the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction, and that can be utilised using the $\Delta\phi(\text{closest jet}, E_{\mathrm{T}}^{\mathrm{miss}})$ distribution. The leading jet p_{T} and H_{T} distributions help to identify a signal enriched, but background depleted region of the phase space. The search also makes use of other two key observables, which are found to be largely uncorrelated:

1. The p_{T} balance between the closest jet (j_1) and farthest jet (j_2) from $E_{\mathrm{T}}^{\mathrm{miss}}$ direction, termed as $p_{\mathrm{T}}^{\mathrm{bal}}$, defined using two-dimensional p_{T} vectors:

$$\Delta_{rel} p_{\mathrm{T}}(j_1, j_2) = \frac{|\vec{p_{\mathrm{T}}}(j_1) + \vec{p_{\mathrm{T}}}(j_2)|}{|\vec{p_{\mathrm{T}}}(j_1| + |\vec{p_{\mathrm{T}}}(j_2)|}$$

   This definition is motivated by double parton interaction measurements, an early example being [227].

2. the difference in the azimuthal angle between j_1 and j_2 as defined above, termed maxminphi, $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$

4.6.1 Inclusive distributions and definition of signal and control regions

Inclusive kinematic distributions are studied for both signal and background processes in order to decide on the analysis strategy. The plots in Fig. 4.2 - 4.5 shows the inclusive kinematic distributions, prior to any kinematic threshold selection. Fig. 4.2 shows the behaviour of signal samples with respect to different r_{inv} fractions, whereas Fig. 4.3 shows the dependence on mediator mass. It was seen that for increasing r_{inv}, the $E_{\text{T}}^{\text{miss}}$ contribution increases as expected, whereas the increasing mediator mass leads to higher hadronic activity due to the effect described in Sec. 2.3.3. It was also observed that no significant change occurs when the dark hadron mass is varied between 10-100 GeV. As can be seen from the Figs. 4.4 - 4.5, the distributions of $p_{\text{T}}^{\text{bal}}$ and $|\phi_{\text{max}} - \phi_{\text{min}}|$ have a significant shape difference between signal and background samples. Similarly, the H_{T} and $E_{\text{T}}^{\text{miss}}$ for the signal is more widespread compared to the SM background. Hence a grid for $E_{\text{T}}^{\text{miss}}$ against H_{T} is constructed, and the $p_{\text{T}}^{\text{bal}}$ and $|\phi_{\text{max}} - \phi_{\text{min}}|$ distributions are looked at in each of these regions.

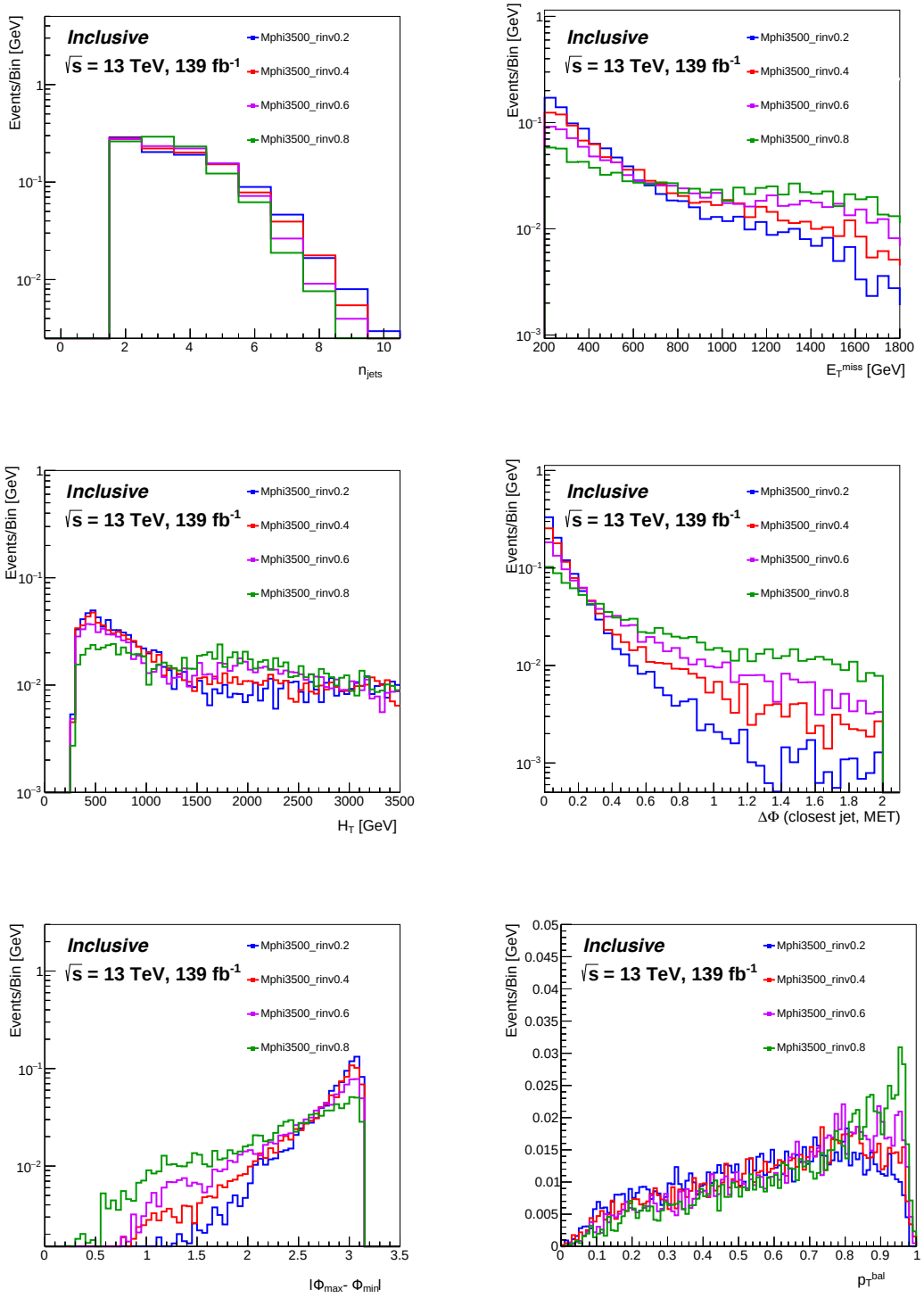


Figure 4.2: Inclusive kinematic distributions of jet multiplicity, $E_{\mathrm{T}}^{\mathrm{miss}}$, scalar jet p_{T} sum, $\Delta\phi$(closest jet, $E_{\mathrm{T}}^{\mathrm{miss}}$), $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$, and $p_{\mathrm{T}}^{\mathrm{bal}}$ for four benchmark signals with different r_{inv} fractions, prior to any kinematic threshold requirement. The signal cross-section is area normalised, and the legend also indicates the mediator mass and r_{inv} fraction.

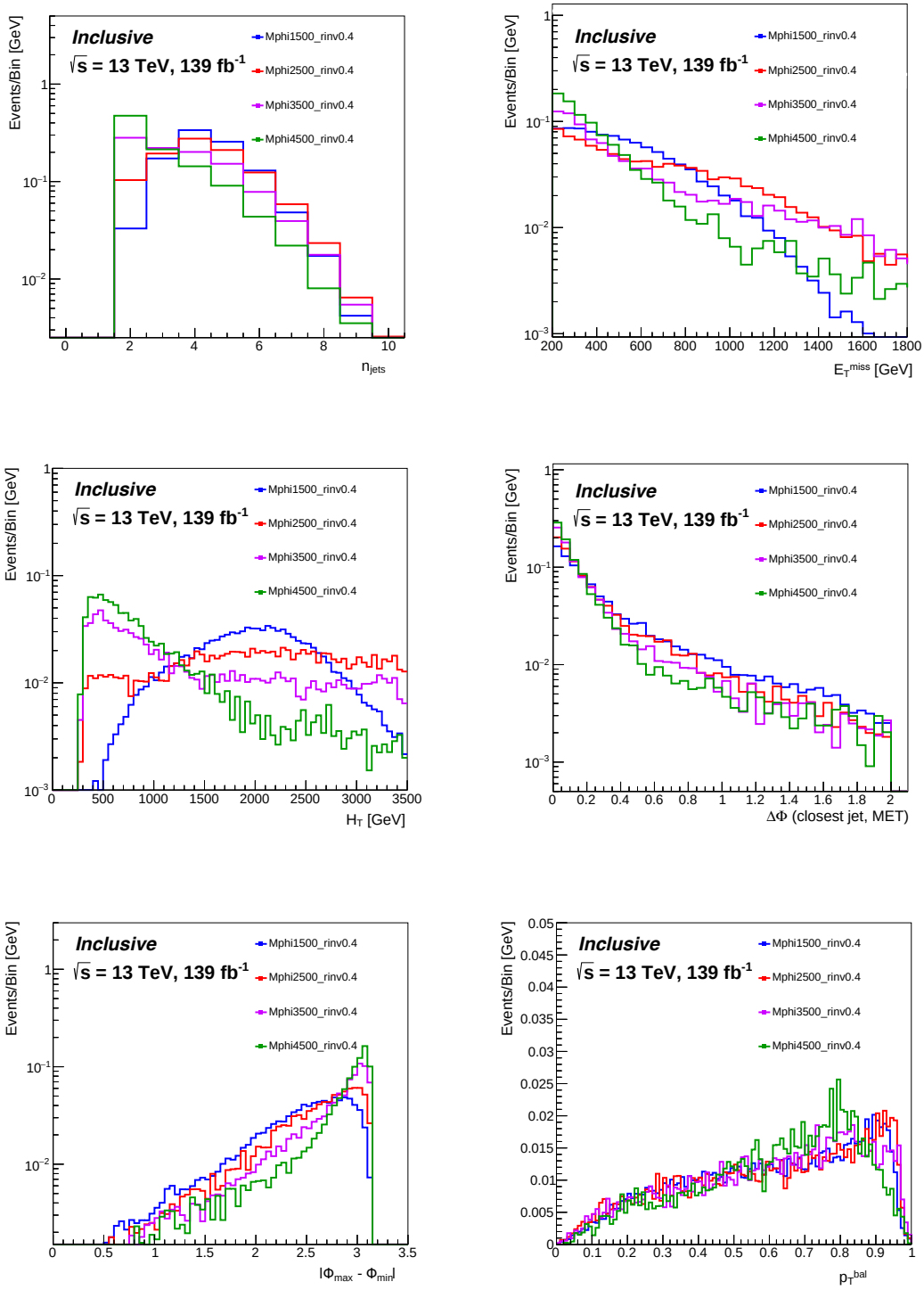


Figure 4.3: Inclusive kinematic distributions of jet multiplicity, $E_{\mathrm{T}}^{\mathrm{miss}}$, scalar jet p_{T} sum, $\Delta\phi$(closest jet, $E_{\mathrm{T}}^{\mathrm{miss}}$), $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$, and $p_{\mathrm{T}}^{\mathrm{bal}}$ for four benchmark signals with different mediator masses, prior to any kinematic threshold requirement. The signal cross-section is area normalise, and the legend also indicates the mediator mass and r_{inv} fraction.

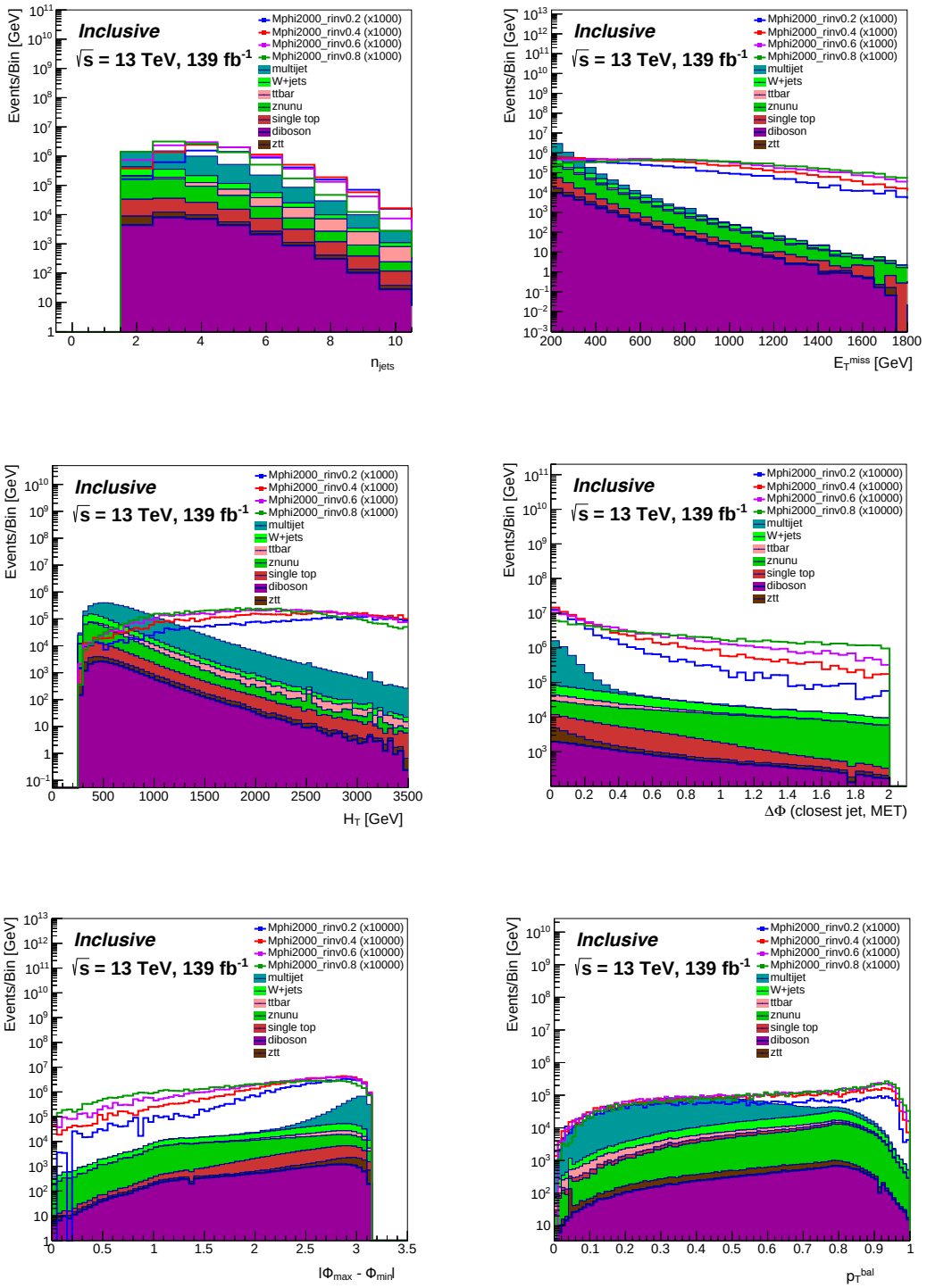


Figure 4.4: Inclusive kinematic distributions of jet multiplicity, $E_{\mathrm{T}}^{\mathrm{miss}}$, scalar jet p_{T} sum, $\Delta\phi(\text{closest jet}, E_{\mathrm{T}}^{\mathrm{miss}})$, $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$, and $p_{\mathrm{T}}^{\mathrm{bal}}$ for four benchmark signal of mediator mass 2000 GeV with different r_{inv} fractions, and background contribution from MC samples, prior to any kinematic threshold requirement. The signal cross-section is scaled as shown in the legend.

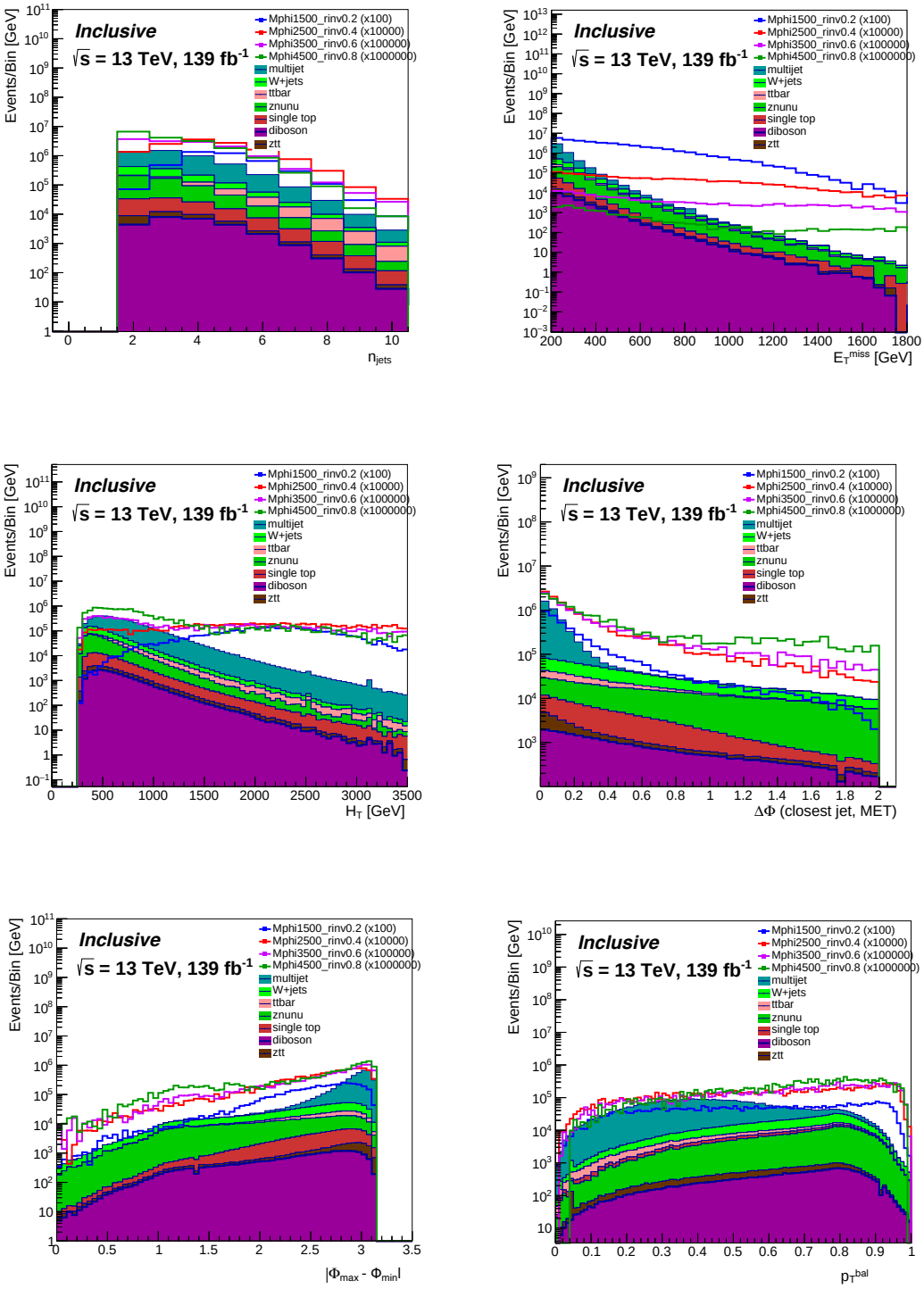


Figure 4.5: Inclusive kinematic distributions of jet multiplicity, $E_{\mathrm{T}}^{\mathrm{miss}}$, scalar jet p_{T} sum, $\Delta\phi$(closest jet, $E_{\mathrm{T}}^{\mathrm{miss}}$), $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$, and $p_{\mathrm{T}}^{\mathrm{bal}}$ for four benchmark signals with different mediator masses and r_{inv} fractions,, prior to any kinematic threshold requirement. The signal cross-section is scaled as shown in the legend.

The Fig. 4.6 shows the correlation between the H_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for a few benchmark signals, multi-jet and the sum of the other SM background processes separately, to check how much they are differently populating the several regions and to show correlation between the two variables. As can be seen from the figure, the two variables seem to be highly uncorrelated.

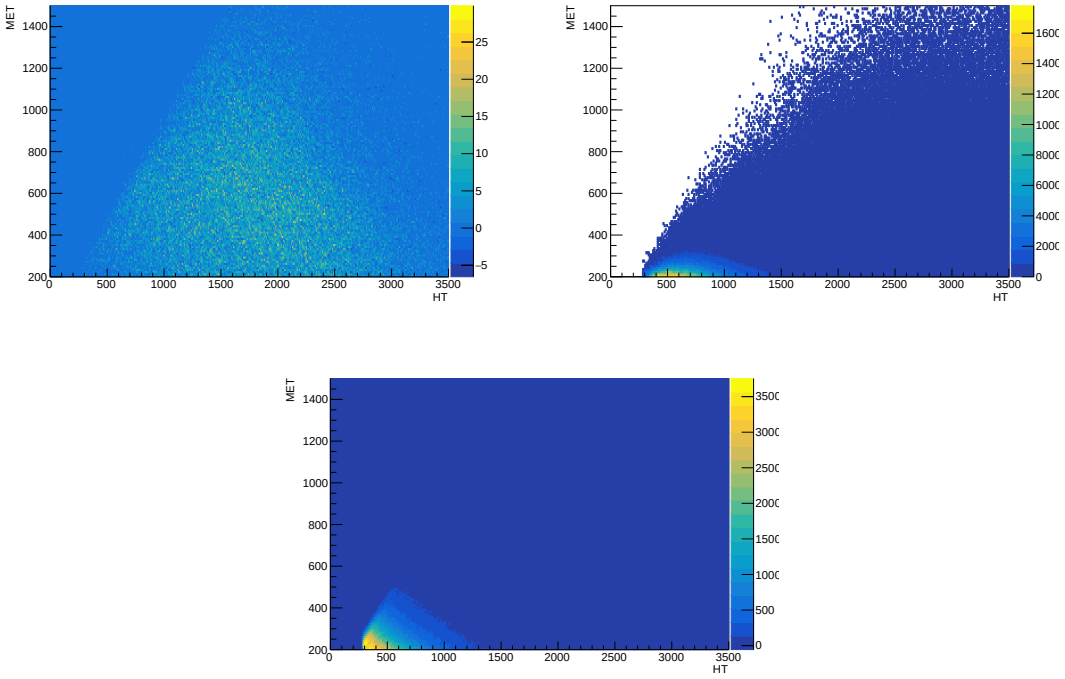


Figure 4.6: Correlation between the H_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for one signal sample with mediator mass of 1.5 TeV and r_{inv} of 0.6, multijet background, and non-multijet background processes summed together.

The following tables 4.10 - 4.13 show the $s/\sqrt{s+b}$ values for the H_{T} optimisation. $H_{\mathrm{T}} > 600$ GeV is picked as the optimal threshold since the significance values for each of these signals are fairly high in that region. It should be noted that the decision to choose $H_{\mathrm{T}} > 600$ GeV is governed by the signal contamination below that threshold and signal significance above that threshold.

H_{T} threshold	Signal (2500, r_{inv}0.4)	Signal (3000, r_{inv}0.4)	Signal (3500, r_{inv}0.4)	Signal (4000, r_{inv}0.4)
< 400	0.032	0.019	0.012	0.007
< 500	0.055	0.043	0.027	0.016
< 600	0.076	0.054	0.032	0.019
< 700	0.089	0.058	0.035	0.021
< 800	0.099	0.061	0.037	0.023
< 900	0.115	0.064	0.039	0.024

Table 4.10: Signal contamination ($s/\sqrt{s+b}$) values for four benchmark signals of different mediator masses in GeV and $r_{\mathrm{inv}} = 0.4$, for determining H_{T} threshold for CR

The signal events typically have high $E_{\mathrm{T}}^{\mathrm{miss}}$, and hence better sensitivity for signals with higher mediator masses and r_{inv} fraction can be expected if the search is performed at a high $E_{\mathrm{T}}^{\mathrm{miss}}$ range. Thus the region with $E_{\mathrm{T}}^{\mathrm{miss}} > 600$ GeV and $H_{\mathrm{T}} > 600$ GeV after the pre-selection is defined as the signal region (SR).

H_T threshold	Signal (2500, r_{inv}0.4)	Signal (3000, r_{inv}0.4)	Signal (3500, r_{inv}0.4)	Signal (4000, r_{inv}0.4)
> 400	1.418	0.364	0.126	0.051
> 500	1.446	0.365	0.123	0.049
> 600	1.514	0.374	0.124	0.047
> 700	1.630	0.397	0.128	0.047
> 800	1.784	0.428	0.135	0.048
> 900	1.972	0.470	0.145	0.050

Table 4.11: Signal significance ($s/\sqrt{s+b}$) values for four benchmark signals of different mediator masses in GeV and $r_{inv} = 0.4$, for determining H_T threshold for SR

H_T threshold	Signal (3500, r_{inv}0.2)	Signal (3500, r_{inv}0.4)	Signal (3500, r_{inv}0.6)	Signal (3500, r_{inv}0.8)
< 400	0.007	0.012	0.012	0.003
< 500	0.019	0.027	0.025	0.009
< 600	0.023	0.032	0.029	0.011
< 700	0.024	0.035	0.032	0.012
< 800	0.026	0.037	0.035	0.012
< 900	0.028	0.039	0.036	0.013

Table 4.12: Signal contamination ($s/\sqrt{s+b}$) values for four benchmark signals of mediator masses in 3500 GeV and different r_{inv} for determining H_T threshold for CR

H_T threshold	Signal (3500, r_{inv}0.2)	Signal (3500, r_{inv}0.4)	Signal (3500, r_{inv}0.6)	Signal (3500, r_{inv}0.8)
> 400	0.088	0.126	0.133	0.039
> 500	0.086	0.123	0.131	0.038
> 600	0.086	0.124	0.133	0.037
> 700	0.089	0.128	0.139	0.038
> 800	0.094	0.135	0.147	0.041
> 900	0.101	0.145	0.158	0.043

Table 4.13: Signal significance ($s/\sqrt{s+b}$) values for four benchmark signals of mediator masses in 3500 GeV and different r_{inv} for determining H_T threshold for SR

Low and intermediate E_T^{miss} validation regions (VR) for multijet process are defined by requiring E_T^{miss} to be between 250 GeV to 300 GeV and between 300 GeV to 600 GeV respectively, with the same $H_T > 600$ GeV requirement after the pre-selection. The signal contamination in the different CRs and VRs are found to be negligible.

4.6.2 Signal region distributions

The signal against total background shape comparison for p_T^{bal} and maxminphi distributions are shown in Fig. 4.8 for SR. There is a distinct shape difference between the different signal benchmark points, and the total background, which is utilised in designing the fit strategy. In order to determine the individual yield of the backgrounds, a simultaneous binned maximum likelihood function fit is performed using the product of all probability distribution functions and nine bin yields, using the MC templates, by employing the SR and the corresponding CRs (1L, 1L1B, and 2L), as will be discussed in detail in Section 4.8. The maxminphi distribution is divided into three bins of size 0–2, 2–2.7, 2.7–3.2, and the p_T^{bal} distribution is divided into three bins of size 0–0.6, 0.6–0.9, 0.9–1, as seen in Fig. 4.9. The same grid is used for the CRs defined subsequently.

The Fig. 4.7 shows the correlation between the p_T^{bal} and maxminphi distributions for signals of different mediator masses and r_{inv} fractions. As can be seen from the figure, the two variables seem to be very less correlated, with an average correlation factor of -0.24.

The tables 4.14 and 4.15 shows the the $s/\sqrt{s+b}$ values for the p_T^{bal} and maxminphi bin

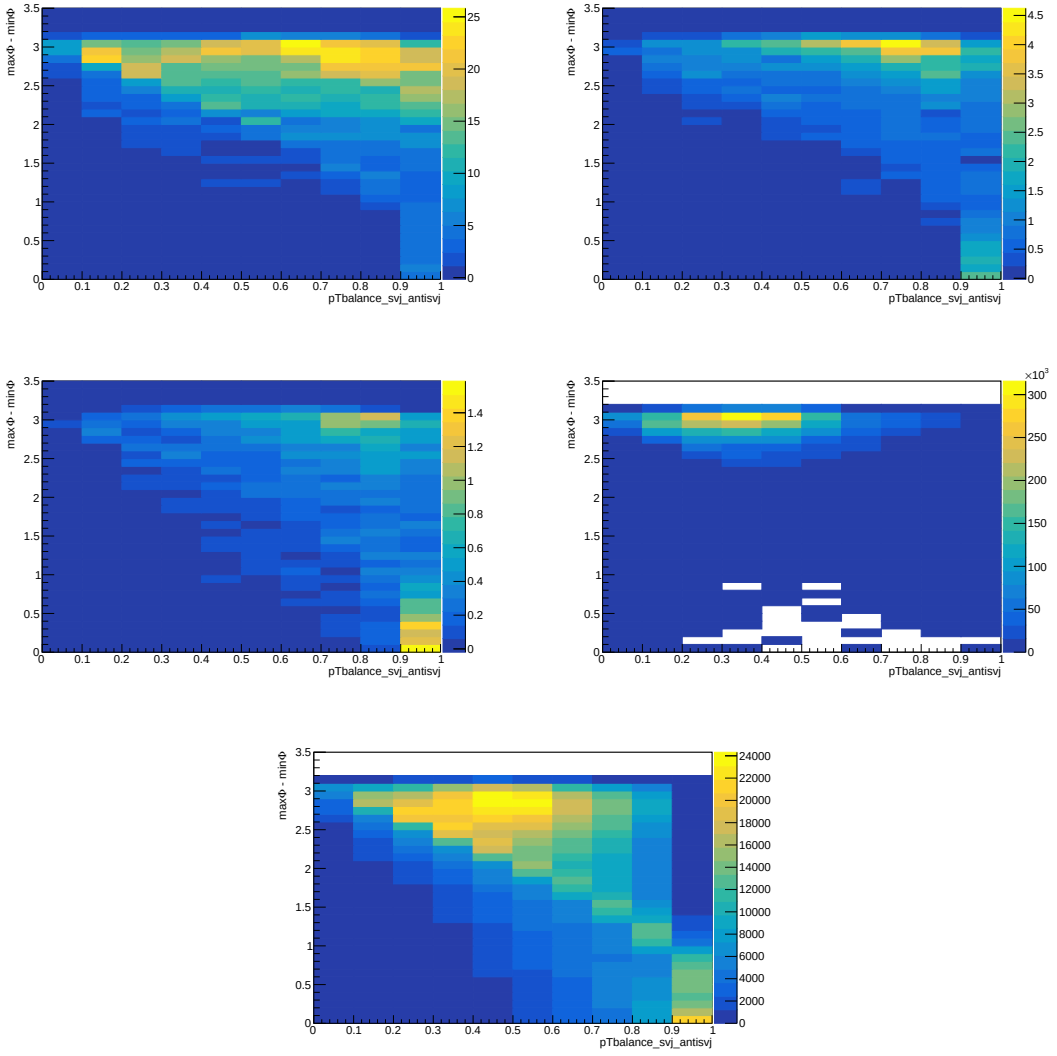


Figure 4.7: Correlation between the p_{T} balance and maxminphi distributions for the different benchmark signal samples with mediator masses 2.5 TeV, 3.5 TeV and 4 TeV with different r_{inv} fractions, multijet background, and non-multijet background processes summed together

optimisation respectively for signal of mediator mass 3.5 TeV and r_{inv} 0.6, to decide the input bins for the final 9 bin grid. Three bins of size 0-2, 2-2.7,2.7-3.2 for maxminphi, and three bins of size 0-0.6, 0.6-0.9, 0.9-1 for the $p_{\mathrm{T}}^{\mathrm{bal}}$ distribution, is picked as the optimal threshold since the significance values for each of these signals are fairly high with that binning.

p_{T} balance ranges	$s/\sqrt{s+b}$
0-0.5, 0.5-0.9, 0.9-1	0.060
0-0.55, 0.55-0.9, 0.9-1	0.061
0-0.6, 0.6-0.9, 0.9-1	0.062
0-0.7, 0.7-0.9, 0.9-1	0.060

Table 4.14: Signal significance ($s/\sqrt{s+b}$) values for signals of mediator mass 3.5 TeV and r_{inv} 0.6, for the $p_{\mathrm{T}}^{\mathrm{bal}}$ bin optimisation for deciding final 9 bin grid.

The following table 4.16 shows the $s/\sqrt{s+b}$ values for the $p_{\mathrm{T}}^{\mathrm{bal}}$ and maxminphi 9 bin grid for signal of mediator mass 3.5 TeV and r_{inv} 0.6. The significance values for the rest of

maxminphi ranges	$s/\sqrt{s+b}$
0-2, 2-2.4 , 2.4-3.5	0.092
0-2, 2-2.5 , 2.5-3.5	0.094
0-2, 2-2.6 , 2.6-3.5	0.095
0-2, 2-2.7 , 2.7-3.5	0.097
0-2, 2-2.8 , 2.8-3.5	0.093
0-2, 2-2.9 , 2.9-3.5	0.091

Table 4.15: Signal significance ($s/\sqrt{s+b}$) values for signals of mediator mass 3.5 TeVand r_{inv}0.6, for the maxminphi bin optimisation for deciding final 9 bin grid.

the signal samples follow a similar trend.

	$p_{\text{T}}^{\text{bal}}< 0.6$	$0.6 < p_{\text{T}}^{\text{bal}}< 0.9$	$0.9 < p_{\text{T}}^{\text{bal}}< 1.0$
maxminphi < 2.0	0.044	0.061	0.120
$2.0 <$ maxminphi < 2.7	0.053	0.076	0.148
$2.7 <$ maxminphi < 3.2	0.035	0.075	0.116

Table 4.16: Signal significance ($s/\sqrt{s+b}$) values for signals of mediator mass 3.5 TeVand r_{inv} 0.6, for for the $p_{\text{T}}^{\text{bal}}$ and maxminphi bin grid used for fitting.

The significance values are defined as $s/\sqrt{s+b}$ and the total is the quadrature sum of the individual bin values. The numbers in this table do not represent the final significance values, but are only used for optimization. The actual significance values are obtained post likelihood fit. While the higher range of these distributions are more signal-enriched, the lower range of maxminphi offers higher signal to background significance. These bins are defined identically in SR and in each CR. Yields in these 9 bins in each case are treated as the observables.

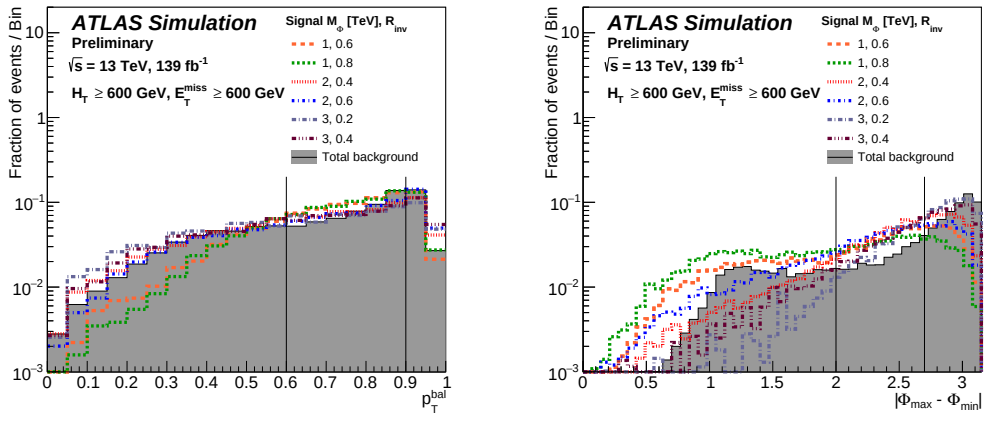


Figure 4.8: Comparisons of shape of $p_{\text{T}}^{\text{bal}}$ and $|\phi_{\text{max}} - \phi_{\text{min}}|$ distributions between the total background and six benchmark signal predictions covering a representative mediator mass and invisible fraction range. The solid vertical lines represent how these distributions are divided to form the 9 bin grid subsequently.

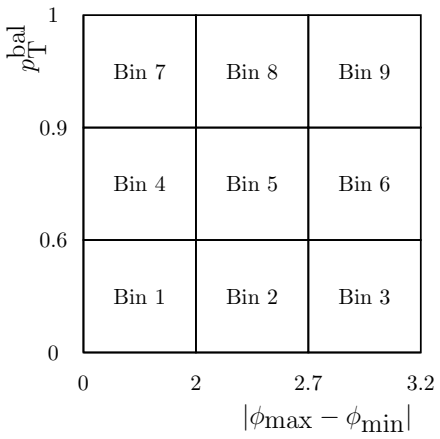


Figure 4.9: The definition of the 9-bins in $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$ and $p_{\mathrm{T}}^{\mathrm{bal}}$, defined identically in SR, VR and in each CR.

4.6.3 Control region distributions

The background contributions are from W/Z+jets, semi-leptonic top processes with real $E_{\mathrm{T}}^{\mathrm{miss}}$, multi-jet processes with mis-measurement of jet momenta and angles, and diboson processes. Fig. 4.10 - 4.12 shows the data against total background comparison for $p_{\mathrm{T}}^{\mathrm{bal}}$ and $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$ distributions for the different leptonic CRs, and similar bins of these two observables are constructed for each of the CRs.

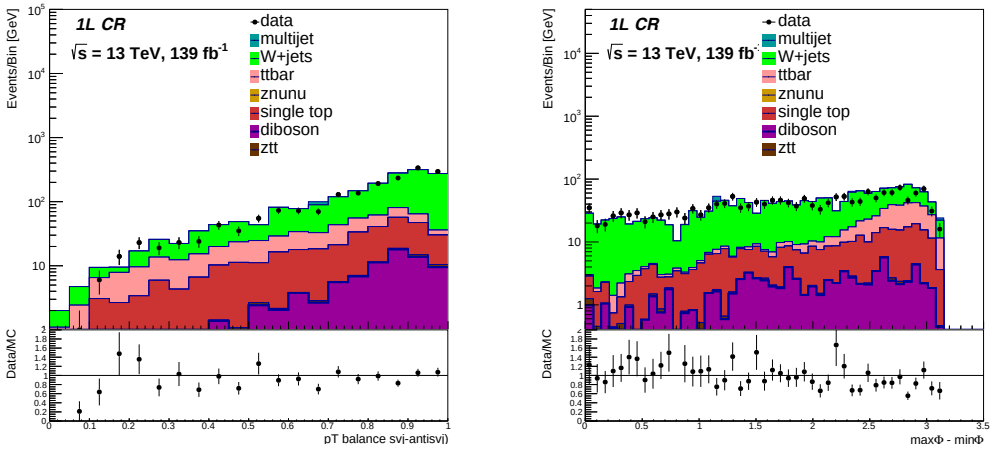


Figure 4.10: Comparisons of shape of $p_{\mathrm{T}}^{\mathrm{bal}}$ and $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$ distributions between the total background and data for 1LCR.

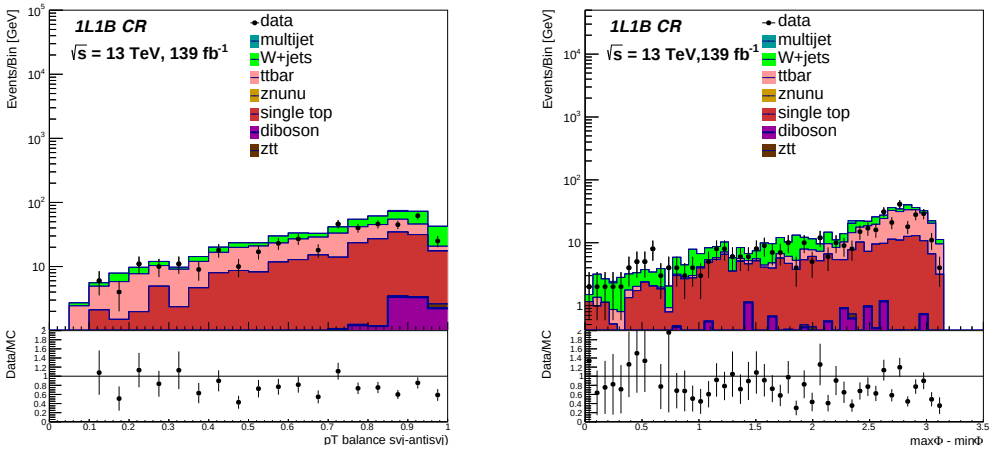


Figure 4.11: Comparisons of shape of $p_{\mathrm{T}}^{\mathrm{bal}}$ and $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$ distributions between the total background and data for 1L1BCR.

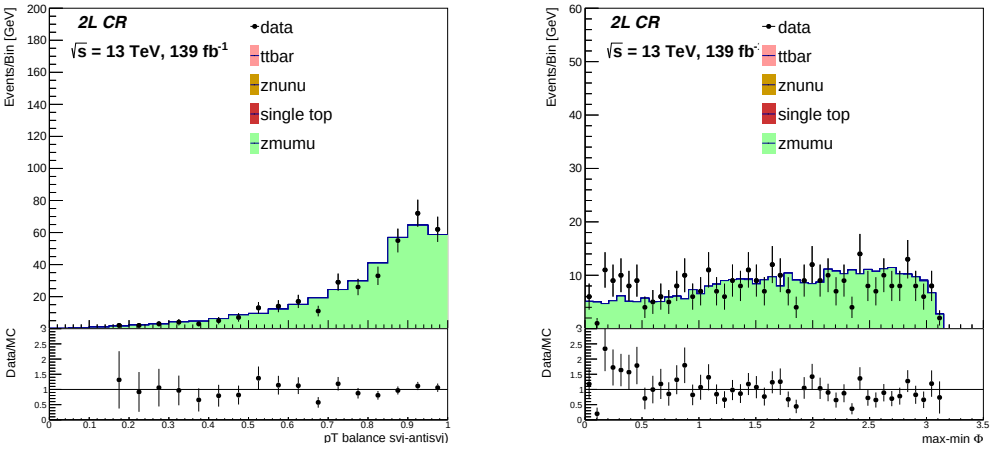


Figure 4.12: Comparisons of shape of $p_{\mathrm{T}}^{\mathrm{bal}}$ and $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$ distributions between the total background and data for 2LCR.

4.7 Systematic uncertainties

The systematic uncertainties on signal and background yields and shapes result from theoretical modelling effects and experimental uncertainties, and all of these sources will be discussed in the following subsections.

4.7.1 Theoretical systematic uncertainties

Theoretical systematic uncertainties affect the MC-based predictions of the electroweak backgrounds to the signal and control regions, and are estimated from comparisons between simulated event yields for different MC generators, as well as from variation of the input parameters used to initialise the event generation. Theoretical uncertainties on the MC samples are due to renormalization and factorization scale, initial and final state radiation and due to PDF choices. The scale and PDF variations (which includes α_S variations as well) are accounted for by individual event weights. The envelop of the variations weights are taken as a combined systematic uncertainty. Also, the available eigentune variations in $t\bar{t}$ and multij-et samples are also considered.

The theory uncertainties for most SM background processes are less than 5% is most bins, except for $t\bar{t}$ processes, where it is up to $20-30\%$ in some bins, but this is observed in many high-p_{T} analysis. The difference across different processes in some bins for PDF or scale variations arise because of limited statistics, or different topology being considered. Typically those bins have very small contributions of those backgrounds to affect the final results.

Treatment of additional systematic uncertainties for $t\bar{t}$ processes

Since $t\bar{t}$ is a dominant background for this analysis, additionally parton shower variation by using an alternate sample with HERWIG7 showering and a matrix element variation by using an alternate sample with mcnlo is considered, as shown in Fig. 4.13. The bin-by-bin ratio of the two samples in the 9-bin SR is taken as an additional systematic uncertainty on $t\bar{t}$ process, since the mismodelling in this process is a general issue in several analyses [228, 229].

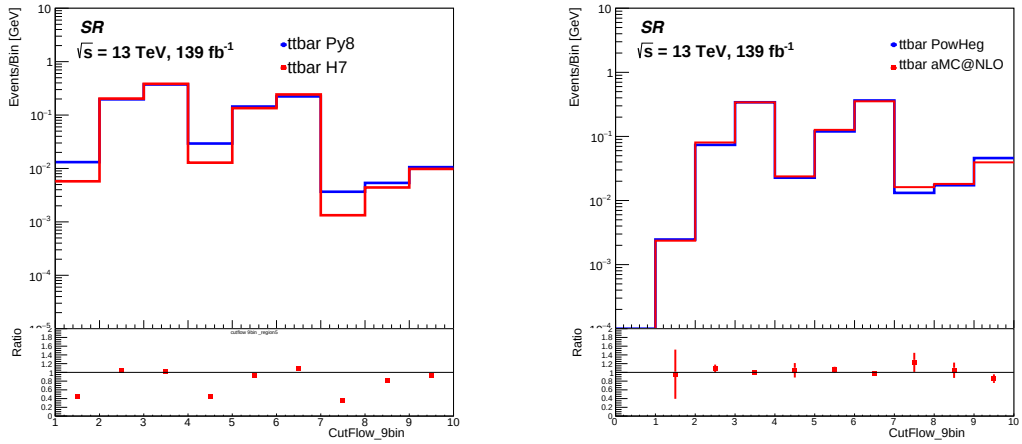


Figure 4.13: Comparison of PYTHIA 8 and HERWIG 7 $t\bar{t}$ samples (left) and comparison of POWHEG 8 and MADGRAPH5_AMC@NLO $t\bar{t}$ samples ME variations (right) in SR. The plots show the area normalised yields for the 9 binned histogram. The bin-by-bin ratio of the two samples in the 9 bin SR will be taken as an additional systematic uncertainty on $t\bar{t}$ process.

Treatment of additional systematic uncertainties for single top processes

Apart from the usual scale, PDF, FSR uncertainties, the comparison between diagram-subtraction (DS) and diagram-removal (DR) schemes [230] to remove the overlap of tW process with $t\bar{t}$ process is also considered, as can be seen in Fig 4.14. The difference is added as an additional uncertainty on tW processes [231].

Treatment of W+jets theoretical uncertainties

The W+jets background in 1L and 1L1B CRs have different composition (enriched in W+light-flavor vs W+heavy-flavor), hence they should be treated separately in the fit. The W+jets theoretical modelling uncertainties are split into W+HF (those with ≥ 1 c or b-jet) and W+LF (the rest), in the simultaneous fit, and uncertainties are keep uncorrelated from each other. Apart from that, since Sherpa 2.2 has been found to underestimate

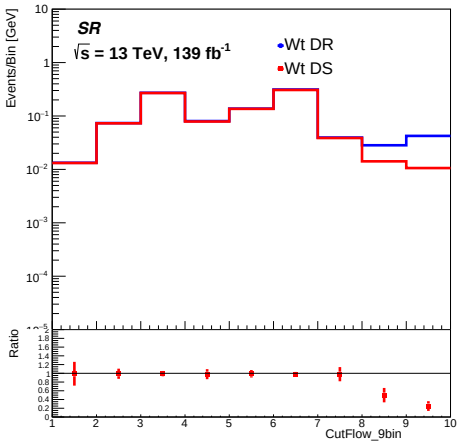


Figure 4.14: Comparison for DR and DS scheme for tW process modelling. The difference is added as an additional uncertainty on tW processes.

W+heavy-flavour by about a factor of 1.3 [232], an additional 30% normalisation uncertainty is assumed for $W+ \geqslant$ 1c+jets and $W+ \geqslant$ 1b+jets subprocesses, and considered correlated between them.

Treatment of Z+jets theoretical uncertainties

The Z+jets scale variation theoretical uncertainty is decorraleted froms bins 1-3, 4-6, and 7-9 in the 9 bin distribution, since in those bins, the differences in $|\phi_{\text{max}} - \phi_{\text{min}}|$ distribution, as can be seen in Fig. 4.12 are captured, for a whole range of p_{T} balance, as shown in the grid in Fig 4.9.

4.7.2 Experimental systematic uncertainties

Experimental systematic uncertainties include uncertainties on energy scales or resolutions of various physics objects, particle reconstruction or identification efficiencies and the integrated luminosity recorded by ATLAS. In the signal region, the dominant uncertainties are those on the jet energy scale and resolution, and due to the lepton veto in the SR, systematic uncertainties on the lepton efficiencies and on lepton energy scale or resolution do not enter the analysis. Efficiency corrections and their associated uncertainties are applied for the use of the Jet Vertex Tagger.

A full list of experimental systematic uncertainties considered is as follows:

Jet energy scale a strongly-reduced nuisance parameter set is used, which combine the effects of the many JES nuisance parameters from underlying variations and make simplifying assumptions about the correlations between these. Jet energies are varied up/down in a correlated manner according to the nuisance parameters [233].

Jet energy resolution uncertainties on the jet resolution evaluated by smearing the jet energies by a Gaussian distribution, whose width is determined from the difference between the jet resolutions measured in MC and data [233].

Jet vertex fraction the efficiencies for tagging pileup or hard scatter jets with the JVT discriminant are corrected in MC by applying scale factors derived by measurement of jet rates in pileup-enriched or depleted phase space regions [198]. Uncertainties on

these scale factors are derived from MC modelling differences and variations in the assumed pileup contamination in the hard-scatter phase space regions.

Missing transverse momentum soft term uncertainties on the hard object components in the $E_{\mathrm{T}}^{\mathrm{miss}}$ are propagated from the systematic uncertainties on the selected objects. Additional uncertainties on the soft $E_{\mathrm{T}}^{\mathrm{miss}}$ component are applied as up/down variations on the scale of the TST, and smearing of the TST by the difference between data and MC resolutions in two orthogonal projections of the soft term with respect to the balancing hard terms.

Pileup-reweighting two variations are applied to the data scale factor used for pileup-reweighting, accounting for uncertainties on the degree of mismodelling of the minimum bias overlay used in MC pileup simulation.

Muon muon reconstruction and identification, momentum scale and resolution, isolation and efficiency of the muon track-to-vertex association uncertainties are accounted for by standard nuisance parameters [234, 235].

τ-lepton the hadronic tau reconstruction and identification efficiency uncertainties are evaluated using the $W \to \tau\nu$ sample, and applied in regions where hadronic τ contributions are expected [208].

Flavour tagging another source of uncertainty related to jets is the one in the scale factor of the b-tagging efficiency It is taken into account in regions where events with b-tagged jets are explicitly selected [202, 204].

Average number of interactions per bunch crossing this is rescaled to improve the agreement of simulation with data, and the corresponding uncertainty, as large as the correction has an effect of 2%

Luminosity the luminosity uncertainty for the full Run 2 dataset of 1.7% is taken into account [236].

4.8 Fit strategy and background estimation

The 9 bin (3 maxminphi bins $\times$ 3 p_{T} balance bins) yields can be treated as the observables for the signals after the $E_{\mathrm{T}}^{\mathrm{miss}}$ and H_{T} selections as illustrated in Fig. 4.9. The contribution of different backgrounds is different for each of the bins, so these 9 bins are used for an estimate of the different background contributions and can be applied to a combined shape fit, as described subsequently.

4.8.1 General strategy and fitting procedure

In order to determine the individual yield of ith background, $N_{\mathrm{i}}^{\mathrm{bg}}$, with a probability distribution function given by P_{i}, a simultaneous binned maximum likelihood function fit is performed using the product of all P_{i} and 9 bin yields, using the MC templates, by employing the SR and the corresponding CRs (1L, 1L1B, and 2L). This is done to simultaneously search for the signal while improving the background prediction in the SR. The scale factors for the individual backgrounds, k_i^{SF} are determined from the fit:

$$\mathcal{L}(\mu,\theta) = \prod_{j\in 36\ \text{bins}} \mathrm{Poisson}(N_j^{\mathrm{obs}}|\mu N_j^{\mathrm{sig}}(\theta) + \sum_{i\in \mathrm{bg}} k_i^{\mathrm{SF}} \times N_{i,j}^{\mathrm{bg}}(\theta)) \times f^{\mathrm{constr}}(\theta) \qquad (4.2)$$

where N_j^{obs} is the observed total yield in the bin j, $N_j^{\text{sig}}(\theta)$ is the total signal yield in in the bin j, and $N_j^{\text{bg}}(\theta)$ is the combined background yield in bin j. The signal strength μ is given by the ratio of the measured to predicted signal cross-section with $\mu \geq 0$. All the systematic uncertainties discussed above are propagated into the simultaneous fit, as different nuisance parameter (NP), denoted by θ. The term $f^{\text{constr}}(\theta)$ represents the product of the Gaussian constraints applied to each of the nuisance parameters, and is defined as,

$$f^{\text{constr}}(\theta) = \prod_{k=1}^{M} G(\theta) \tag{4.3}$$

where G is the standard normal distribution, M is the total number of systematic uncertainty sources, and θ represents the up- and down-variation yields. The MC templates are allowed to vary within their shape uncertainty, and all the systematic uncertainties discussed above are propagated into the simultaneous fit. The NPs are correlated across bins and signal and control regions unless stated otherwise. The fit finds the best set of values for the unknown parameters μ and θ that maximizes $\mathcal{L}$ and the combined uncertainty on μ taking into account the correlations between all sources of uncertainty.

The test statistics are defined as the likelihood ratio:

$$\lambda(\mu) = -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}. \tag{4.4}$$

Upper limits on the contribution of events from new physics are computed by using the modified frequentist approach CL_s based on asymptotic formulas at 95% confidence level [237].

Fitting to Asimov data To avoid biasing analysis results, the data is not used in the fit until unblinding. The simulated dataset is used as a representation of the expected dataset in the fit to validate the behavior of the fit setup and extract expected sensitivity. This dataset is referred to as the *Asimov Dataset*. Using the Asimov dataset in the fit fixes the signal strength parameter, $\mu = 1$, and all nuisance parameters take on their nominal values (θ_j). The covariance matrix is calculated along with the expected uncertainty on μ and the relative impacts of each systematic uncertainty.

4.8.2 Verification of multijet background estimation

Since this analysis final state involves jets and $E_{\text{T}}^{\text{miss}}$, the multijet background with the high $E_{\text{T}}^{\text{miss}}$ selection arises predominantly because of mis-measurement of jet momenta and angles, thereby disrupting the actual well-balanced configuration. It can also arise because part of the jet ending up in a dead calorimeter cell. While the dead tile veto is expected to take care of the last scenario, the first case requires a dedicated data driven method for cross-checks.

In order to verify the behaviour of multijet background in the SR, a comparison of data with background is done in a VR with moderate $E_{\text{T}}^{\text{miss}}$ ($250 < E_{\text{T}}^{\text{miss}} < 300$ GeV), and $H_{\text{T}} > 600$ GeV. Different kinematic distributions are plotted in this region to see how the multijet process behaves, before fitting. These distributions are shown in Fig 4.15. If there was no mismodelling, the data-otherMC term should overlap with the multijet MC, however, that is not the case. Here otherMC denotes the non multijet MC contributions. There is known mismodelling in multijet, so a data-otherMC vs multijet reweighting is done in in the 9 bin distribution in this region, where the reweighting factors are obtained in bin 3, 6, 9, and applied to 1-3, 4-6, 7-9 respectively. This approach is done because most mj resides

in bins 3, 6, and 9, so they can be treated as normalisation baselines for the rest of the pt-balance ranges. Since a reweighting factor cannot be derived reliably for each of the bins, the reweighting factors are obtained just in the well measurde bins, i.e. bin 3, 6 and 9. Those reweighting factors are applied to the rest of the bins in the same $p_{\mathrm{T}}^{\mathrm{bal}}$range.

Treatment of additional systematic uncertainties for multijet processes

The MJ normalization is varying freely in the likelihood fit and adjusted to data. Additional uncorrelated shape uncertainties are considered in the multijet template. In particular a 25% normalization uncertainty is assigned to bins 6 and 9. This uncertainty is increased to 50% in all intermediate maxminphi bins (bins 2, 5, and 8) and to 100% in the lowest maxminphi bins (bins 1, 4, and 7). These uncertainties have a only a small impact on the sensitivity of the search. Increasing them by a factor of two would only degrade the expected signal cross section limit by 2%.

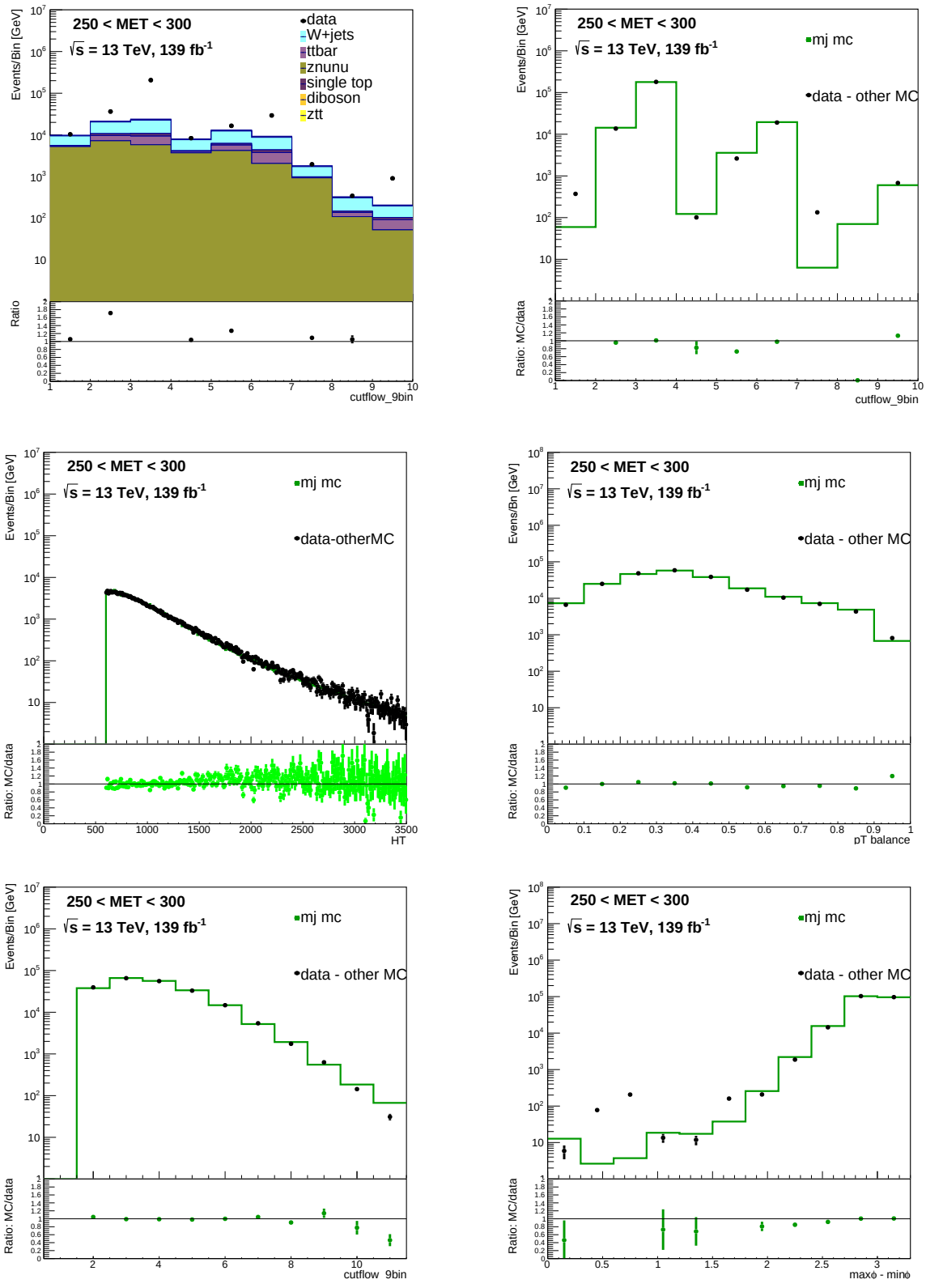


Figure 4.15: Kinematic distributions of the 9 bin yields, scalar jet p_{T} sum, $p_{\mathrm{T}}^{\mathrm{bal}}$and maxminphi for estimation of multijet background, using H_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ VR.

4.9 Simulaneous fit of signal region and control region

The background estimation strategy must be robust, and hence the above mentioned leptonic CRs are included in the fit.

4.9.1 Post-fit distributions in 1LCR

The same 9-bin grid of p_{T} balance and maxminphi distributions, described in Fig. 4.9 are studied in the different $E_{\mathrm{T}}^{\mathrm{miss}} > 600$ GeV vs $H_{\mathrm{T}} > 600$ GeV region, and this is chosen as the primary W+jets processes control region. The formula described in Eq. 4.2 is utilised, and a 9-bin fit of W+jets processes background is performed, by setting the other background yields, apart from multijet, to be free floating. The multijet yield is set to zero for this CR, since we do not expect leptonic contributions from multijet.

Fig. 4.16 shows the fitted histogram in 1LCR. Here, the diboson background scale factor is set to be fixed to 1. The fit shows agreement between data and MC in most bins.

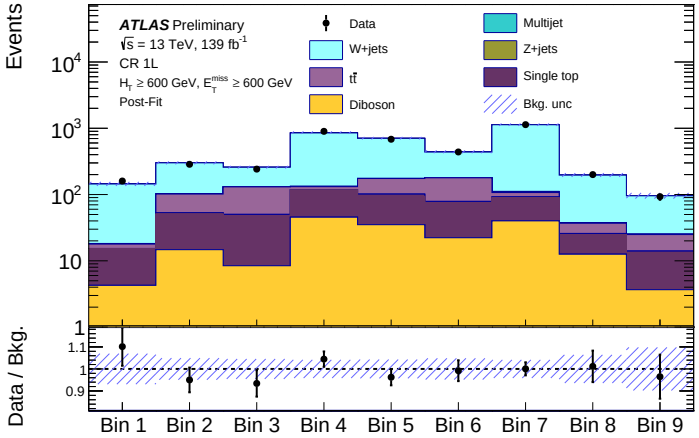


Figure 4.16: 9 bin fitted histogram in $E_{\mathrm{T}}^{\mathrm{miss}}$ vs H_{T} in 1LCR. This is the primary W+jets processes control region. The uncertainty band includes all experimental and theoretical systematic uncertainties, including MC statistical uncertainty.

4.9.2 Post-fit distributions in 1L1BCR

The 9-bin grid of p_{T} balance and maxminphi distributions, described in Fig. 4.9 are studied in the different $E_{\mathrm{T}}^{\mathrm{miss}} > 600$ GeV vs $H_{\mathrm{T}} > 600$ GeV region, and this is chosen as the primary top processes estimation region. The formula described in Eqn. 4.2 is utilised, and a 9-bin fit of top processes background is performed, by setting the other background yields, apart from multijet, to be free floating. The multijet yield is set to zero for this CR, since we do not expect leptonic contributions from multijet

Fig. 4.17 shows the fitted histogram in 1L1B CR. Here, the diboson background scale factor is set to be fixed to 1. The fit shows agreement between data and MC in most bins.

4.9.3 Post-fit distributions in 2L CR

Finally, for the Z+jets processes estimation, a 2L selection is employed as discussed in Sec 4.5, and this selection ensures a pure $Z \to \mu\mu$ sample. The two opposite sign muons are treated as neutrinos, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is recalculated accordingly. To eliminate contributions

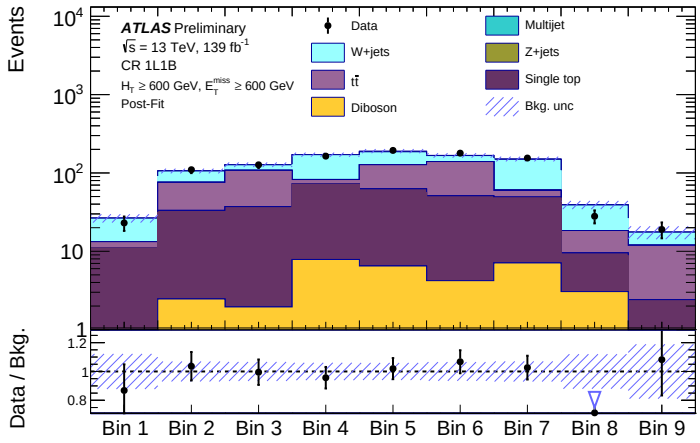


Figure 4.17: 9 bin fitted histogram in $E_{\mathrm{T}}^{\mathrm{miss}}$ vs H_{T} in 1L1BCR. This is the primary top processes estimation region. The uncertainty band includes all experimental and theoretical systematic uncertainties, including MC statistical uncertainty.

from top processes, an additional requirement is imposed, which demands the events to have no b-tagged jets. This is because $Z \to \nu\nu$ and $Z \to \mu\mu$ should have same kinematics, and the additional jets yield events with $E_{\mathrm{T}}^{\mathrm{miss}}$ close to a jet. Following the event selections (described in Sec. 4.5) for the 2L control region, a similar $E_{\mathrm{T}}^{\mathrm{miss}}$ vs H_{T} threshold is employed for Z+jets backgrounds in this case. The following plots in Fig. 4.18 shows the fitted histogram, passing all kinematic thresholds, for the $E_{\mathrm{T}}^{\mathrm{miss}}$ vs H_{T} in 2L CR. The event selection requirements for this region ensures a pure $Z \to \mu\mu$ sample, as can be seen from the figure below, and the fit shows agreement between data and MC in most bins.

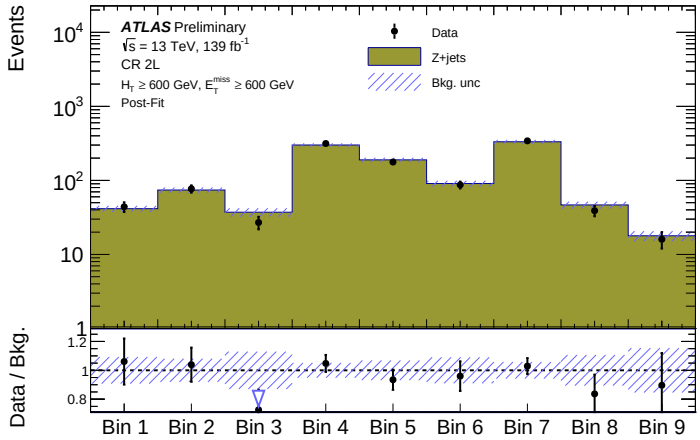


Figure 4.18: 9 bin fitted histogram in $E_{\mathrm{T}}^{\mathrm{miss}}$ vs H_{T} in 2LCR. This is the primary Z+jets processes control region.

4.9.4 Post-fit distribution for combined fit in control region and signal region

The Fig 4.19 shows the unblinded 9-binned distributions with uncertainty bands having all the systematic uncertainties included, whereas Fig 4.20 shows the ranking of the dominant systematic uncertainty sources for SR. As can be seen, the different NPs behave well within

the one standard deviation threshold, and the correlation matrix between the different systematic uncertainty sources are shown in Fig 4.21. Additionally, Table 4.17 provides the total yield for each background process and a few benchmark signal points.

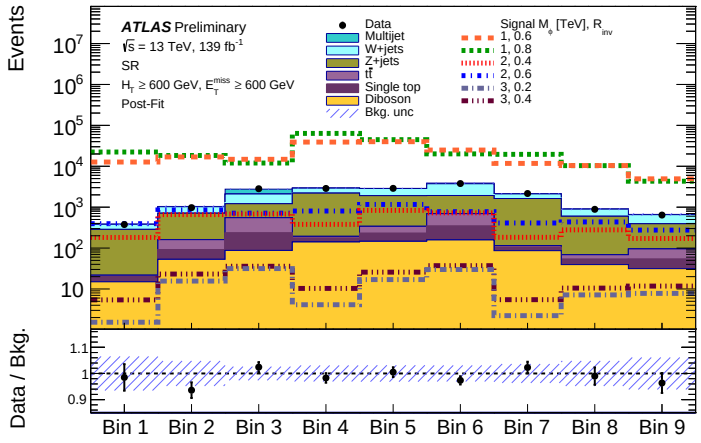


Figure 4.19: Postfit 9 bin distribution for unblinded SR, with systematic uncertainties, BG only fit, overlaid for a few representative signal mass points.

	SR	CR 1L	CR 1L1B	CR2L
Z+jets	8490 ± 260	11.6 ± 1.4	2.2 ± 0.6	1120 ± 40
W+jets	5820 ± 300	3190 ± 170	351 ± 41	-
$t\bar{t}$	920 ± 70	350 ± 29	304 ± 24	-
Single top	533 ± 47	358 ± 29	290 ± 25	-
Multijet	850 ± 100	28 ± 11	7.7 ± 3.1	-
Diboson	757±10	187 ± 9	34.5 ± 2.8	-
Total background	17370 ± 280	4120 ± 100	990 ± 35	1120 ± 40
Data	17388	4136	999	1124
Signal:				
M_ϕ= 1 TeV, r_{inv}= 0.6	180000 ± 40000	-	-	-
M_ϕ= 1 TeV, r_{inv}= 0.8	220000 ± 50000	-	-	-
M_ϕ= 2 TeV, r_{inv}= 0.4	4100 ± 900	-	-	-
M_ϕ= 2 TeV, r_{inv}= 0.6	5800 ± 1300	-	-	-
M_ϕ= 3 TeV, r_{inv}= 0.2	117 ± 26	-	-	-
M_ϕ= 3 TeV, r_{inv}= 0.4	170 ± 40	-	-	-

Table 4.17: Post-fit yields from background-only fit, including pre-fit contributions of different signal benchmark points. Dashes refer to components that are negligible or not applicable.

Table 4.18 shows the fitted normalisation factors for the different background processes. As can be seen, after the data-otherMC reweighting, the multijet process normalisation factor is close to unity.

Process	k_i^{SF}
Z+jets	1.18 ± 0.05
W+jets	1.09 ± 0.04
Top processes	0.64 ± 0.04
Multijet	1.10 ± 0.04

Table 4.18: Scale factors for each background processes obtained from simultaneous fit using SR, 1L CR, 1L1B CR and 2L CR. Top processes denotes merged contributions from $t\bar{t}$ and single top processes.

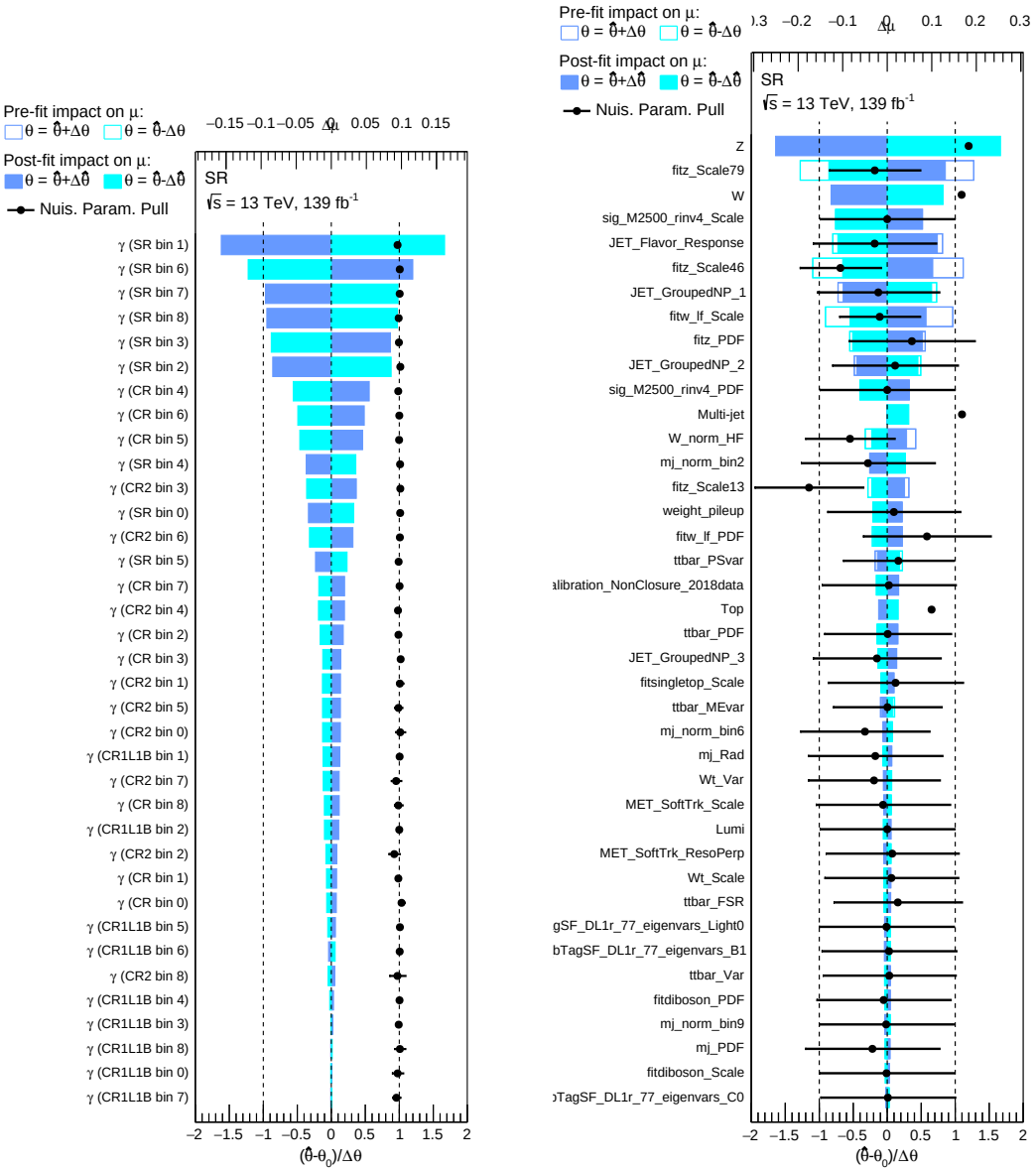


Figure 4.20: Ranking of nuisance parameters for CR-SR simultaneous fit, showing the dominant contributions from systematic uncertainty sources and the MC statistical uncertainty. The signal of $M_\phi = 2.5$ TeV, $r_{\text{inv}} = 0.4$ is used here. Nuisance parameters are ranked according to their impact on the observed signal strength, which is evaluated by fixing each respective nuisance parameter to a value shifted by its uncertainty, repeating the fit, and comparing the signal strength to the nominal fit result. Empty boxes with blue borders indicate pre-fit impacts, whereas post-fit impacts are shown as filled boxes. Pulls of the nuisance parameters and their post-fit uncertainties are shown as black points and error bars. The γs are the statistical uncertainty in each bin of the CRs and SR.

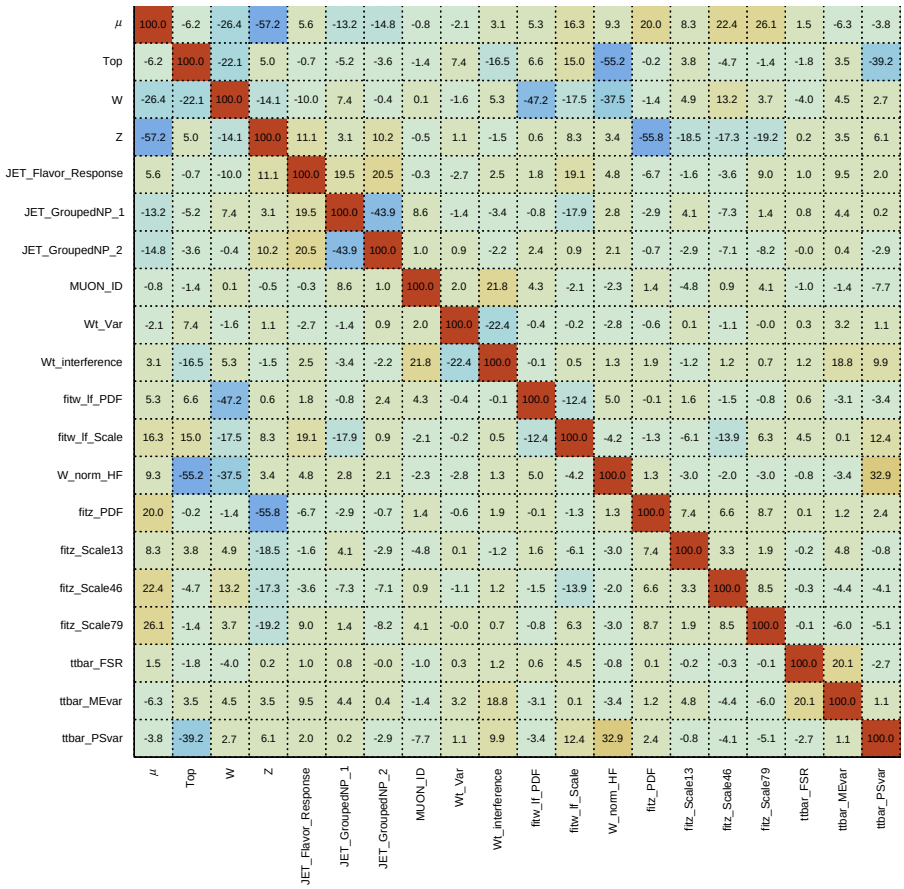


Figure 4.21: Correlation matrix for CR-SR combined fit, with a representative signal of mediator mass of 2.5 TeVand r_{inv} 0.4. This shows the percentage correlation between all the systematic uncertainty sources that are contributing to the simultaneous CR-SR fit.

4.10 Results

The final results are presented in terms of different kinematic distributions and exclusion limits, since no significant excess was observed beyond SM predictions. Fig 4.22 shows the kinematic variables for data vs background comparison, with a few representative signal mass points overlaid for the SR. All figures have uncertainty bands with all the systematic uncertainties included. Again, excellent agreement of data with SM background predictions are seen for all the observables.

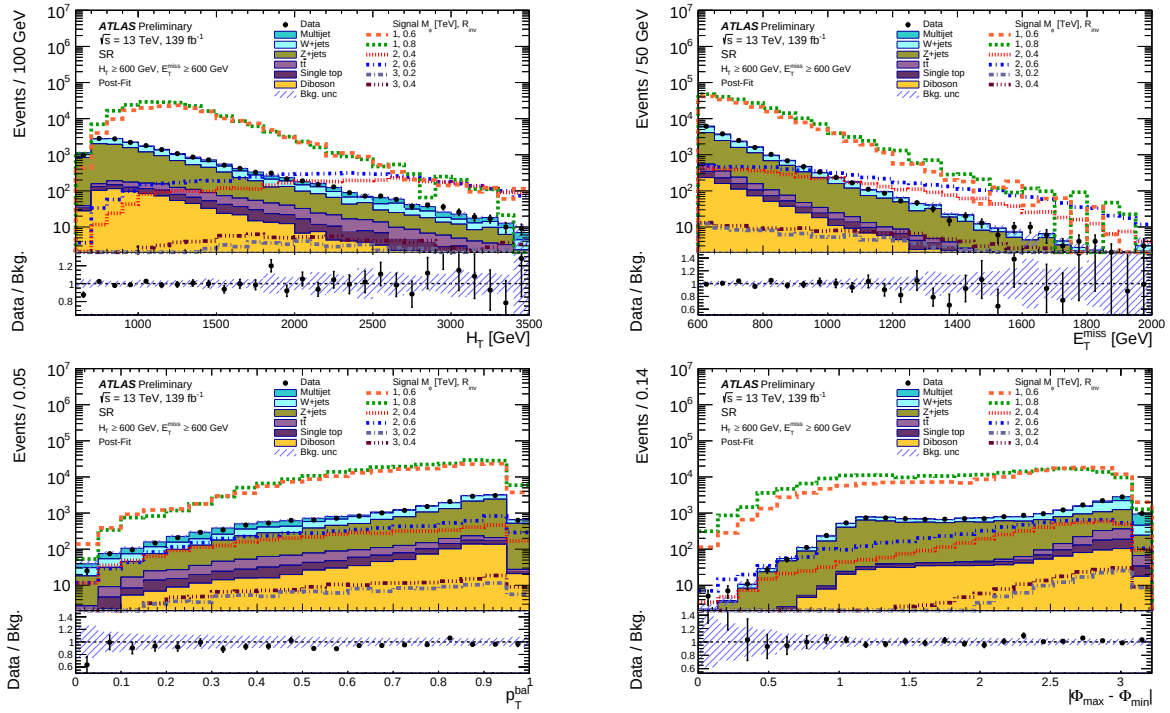


Figure 4.22: Postfit distributions of unblinded SR kinematic variables for scallar jet p_{T} sum, $E_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\mathrm{bal}}$and maxminphi, with all systematic uncertainties, overlaid for a few benchmark signal mass points.

Upper limits on the contribution of events from new physics are computed by using the modified frequentist approach CL_s based on asymptotic formulas at 95% confidence level [237]. The compatibility of the observed data with the different signal predictions is estimated by the limit on the signal strength, μ, defined previously. For values of $\mu \leq 1$, the nominal cross-section is excluded, while for $\mu > 1$, no such conclusion can be obtained. The limits on the signal model are presented in two different ways. The 95% exclusion limit plots for limits on cross-section as a function of mediator mass is shown for each r_{inv} value in Fig. 4.23. The observed limits increase from 2.4 TeV for r_{inv} of 0.2, to 2.7 TeV for r_{inv} of 0.8. The limits degrade at higher masses as the shapes of the discriminating observables such as $|\phi_{\mathrm{max}} - \phi_{\mathrm{min}}|$ become more background-like. The absence of systematic uncertainties would have improved the limits by about 25%.

Additionally, the nominal signal cross-sections for each signal mass point can be scaled by λ^4 for mediator masses over 2.5 TeV without any change in kinematic distributions. For each mediator mass point, the limit on the cross-section is obtained, and the corresponding λ is calculated, using $\mu^{1/4}$. This λ value corresponding to the cross-section upper limit is presented for SR in Fig. 4.24.

It can be seen that for lower mass points, the nominal cross-sections are excluded, whereas

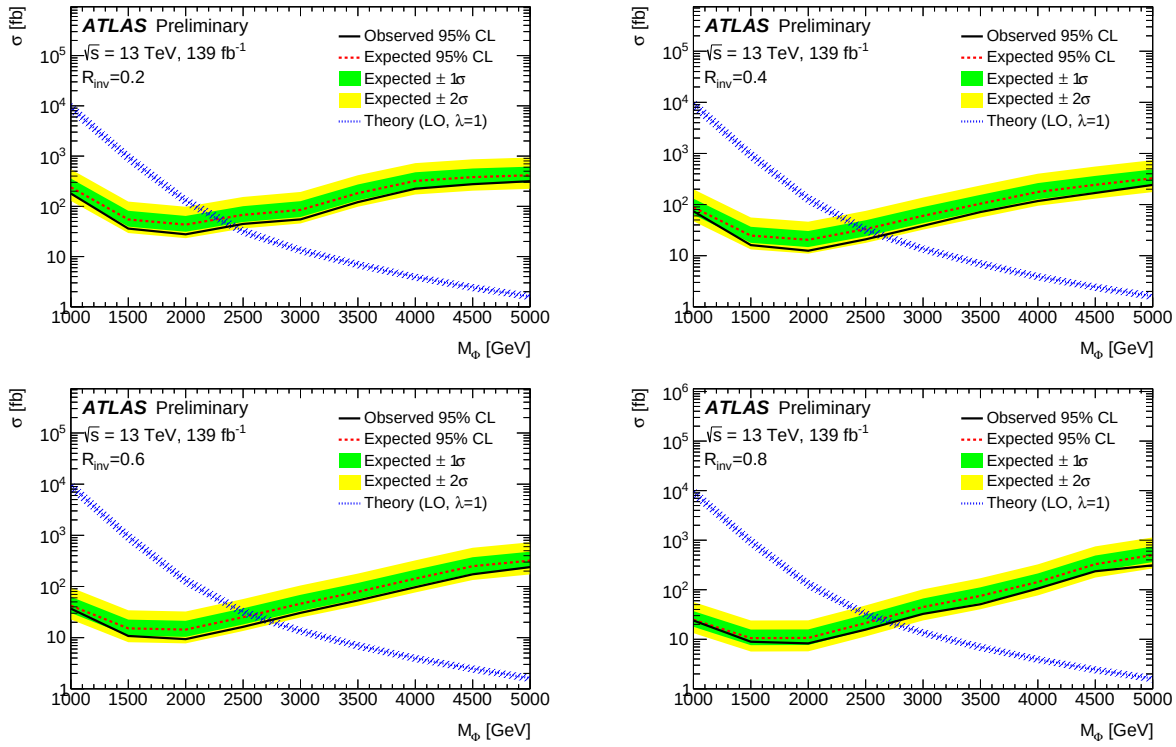


Figure 4.23: The 95% CL upper limit on the semi-visible jet production cross-section as a function of mediator mass are shown for invisible fraction of 0.2, 0.4, 0.6, and 0.8. The green and yellow shaded bands correspond to expected one and two standard deviation uncertainty respectively, with the expected central value shown by the red dashed line. The solid black line is observed limit as a function of the mediator mass. The LO theory prediction (for $\lambda = 1$) with the uncertainty band is shown by the blue line.

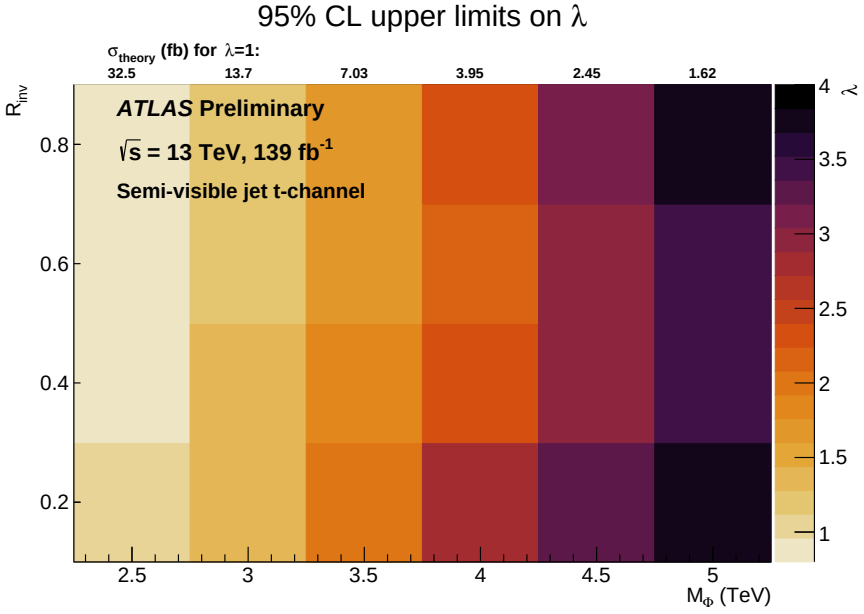


Figure 4.24: The grid shows the observed 95% CL upper limit on λ with M_ϕ on the x-axis, r_{inv} on the y axis. It also includes over each M_ϕ column the predicted cross-section for that specific mass value as a reference.

for higher mass points only higher values of cross-sections can be excluded. The advantage of this representation is that it sets stringent limits on the signature in general for a wide range of λ values, and can help in recasting this analysis for future model predictions. The absence of systematic uncertainties would have improved the limits by 10 - 20%. The largest contribution to systematic uncertainties arise from signal modelling uncertainties, followed by theoretical uncertainties on other background processes.

4.11 Conclusions

A search for semi-visible jets in the t-channel production mode in pp collisions for an integrated luminosity of 139 fb^{-1} at $\sqrt{s} = 13$ TeV at the LHC, based on data collected by the ATLAS detector during 2015-2018 was presented in this chapter. The first limits on the SVJ t-channel production for mediator masses ranging from 1000–5000 GeV, and for r_{inv} of 0.2–0.8 have been set by this analysis, since no significant excess was observed beyond SM predictions. The observed yields are in agreement with the SM background expectations. The upper limits at the 95% confidence level on the mediator mass range from 2.4 TeV to 2.7 TeV, depending on the values of the invisible energy fraction. They are translated into upper limits on the coupling strength between the mediator, a Standard Model quark and a dark quark.

Chapter 5

Phenomenological explorations of semi-visible jets

While searches for semi-visible jets (SVJ) are underway in the LHC experiments, with the first results from ATLAS [195] and CMS [238] experiments, the viability of such searches in other similar topologies with or without alternate definition of jets are probed in parallel. This chapter will focus on three exploratory studies performed to enhance the sensitivity of SVJ detection. The more challenging scenario of t-channel production mode of semi-visible jets is looked at, where the absence of a resonance mass peak makes identifying the distinct signal features more critical. The two-vertex coupling strength, λ as mentioned in the previous chapter, can be treated as a free parameter to gauge the sensitivity of the SVJ signal with respect to the background. For these studies, the Rivet [239] analysis toolkit was used, with the Fastjet package [240] for jet clustering.

5.1 Event generation

The signal samples, at $\sqrt{s} = 13$ TeV are generated by using a t-channel simplified dark-matter model in MadGraph [209] matrix element (ME) generator, with $xqcut = 100$ [1] and NNPDF2.3 LO PDF set [241], for a mediator mass of 1500 GeV and a dark-matter candidate mass of 10 GeV. Different r_{inv} fractions result in somewhat different kinematics, so r_{inv} values of 0.25, 0.50 and 0.75 are studied, as well as the values of zero (no dark component) and unity (fully dark jet) corresponding to the boundary conditions in Sec 5.2. r_{inv} values of 0.3, and 0.7 are explored in Sec 5.3 for the identification of new observables. The process $pp \to \chi\bar{\chi}$ with up to two extra jets were simulated and MLM matched [63] in order to have a reasonable cross-section and obtain a proper signal which does not get swamped under multijet background. The multijet production described by QCD are generated with Pythia 8. As the substructure of the jets is of primary interest, the lack of higher order matrix element in background simulation is not a concern. Multijet events were generated using Pythia 8 as before [64]. No dramatic difference is observed in substructure observables. Pythia 8 model does take into account the effect of heavy flavour jets created by gluon splitting. It must be noted though, that while multijet events at particle level mostly have low values of missing transverse momentum, at detector level, due to mis-measurement of jet energy and direction, a large fraction of events acquire large values of missing transverse momentum.

[1] defined as the minimum k_{T} separation between partons

5.2 Exploring jet substructure of semi-visible jets

This section is primarily based on Ref [242], unless stated otherwise.

5.2.1 Analysis strategy

In this study, large-radius jets are considered, more specifically anti-k_t [171] jets with R=1.0, trimmed (with $R_{sub} = 0.2$ and $f_{cut} = 0.05$) [243] in order to stay close to potential experimental analysis. The validity of this can be seen in Fig. 5.1, where it is evident that large-radius jets better contain the semi-visible jets. Reclustered jets [200] may also be a good way to probe these events, but that has been left for a future study.

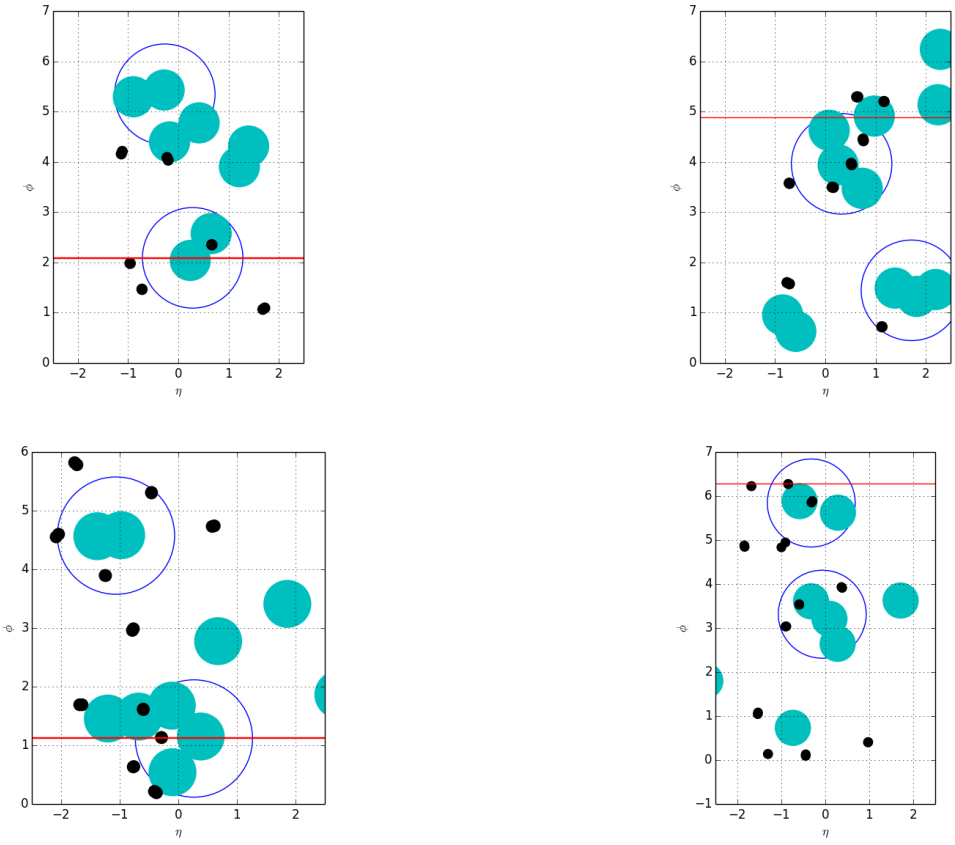


Figure 5.1: Various objects are plotted in $\eta - \phi$ plane for four representative signal events. The large hollow blue circles represent the trimmed large-radius jets, the filled cyan circles represent anti-k_t jets with R=0.4, the black points represent dark hadrons, and the red line the direction of missing transverses momentum.

The large-radius jets are required to have $p_T > 250$ GeV. As stated previously, the identifying signature of SVJ is the alignment of the event missing transverse momentum along the direction of such a jet. Therefore it is required that there be at least one large-radius jet within $\Delta\phi < 1.0$ of the missing transverse momentum direction, and that jet is tagged as a SVJ. Additionally, missing transverse momentum is required to be at least 200 GeV, owing to the fact that an actual search using a missing transverse energy trigger will require that threshold.

It is however interesting to note that in a majority of events, the subleading jet in transverse momentum is tagged as the SVJ, as can be see from the distribution of $\Delta\phi$ between leading and subleading jets with the missing transverse momentum direction in Fig. 5.2.

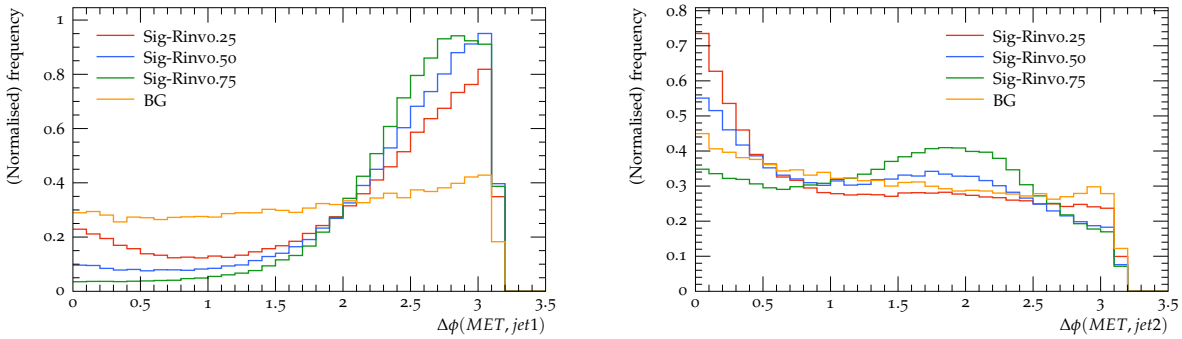


Figure 5.2: Distributions of the azimuthal angle difference between the leading (left) and subleading (right) jets with the direction of missing transverse momentum for three different signals corresponding to r_{inv} values of 0.25, 0.50 and 0.75, and the background.

In Fig. 5.3, it is shown that the events with SVJ have high missing transverse momentum compared to the background jets, as expected, and also the p_{T} distribution of the SVJ with the background jets. The leading large-radius jet is selected, without any requirement on missing transverse momentum as the background jet. It should be noted that even though the SVJ is more often than not the sub-leading jet, the main interest is in differentiating SVJ from standard quark/gluon-initiated jets, so the leading jet can be used from the background without any loss of generality. It was observed that using only quark or only gluon initiated background jets made no difference.

Apart from the multijet process, which is the dominant background, W/Z + jets processes can contribute to the background. However, the processes with one or more leptons can be almost completely rejected by vetoing events with leptons.

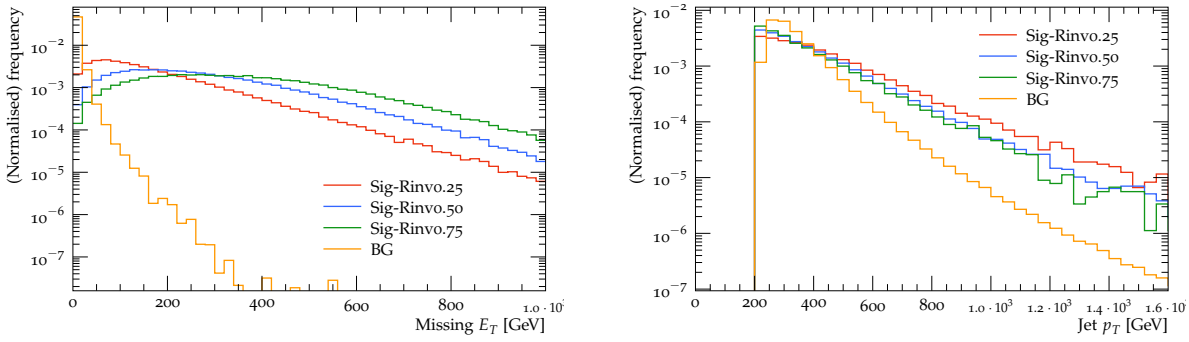


Figure 5.3: Distributions of $E_{\text{T}}^{\text{miss}}$ (right) and leading jet p_{T} (left) for three different signals corresponding to r_{inv} values of 0.25, 0.50 and 0.75, and the background.

5.2.2 Jet substructure observables

Many jet substructure observables have been designed over last decade or so [244, 245], with different sensitivity to different signal jets. In recent works, the focus was on energy correlation observables [246], and discussed the non trivial theoretical uncertainties associated

with jet substructure. In this study, a broad array of observables were explored, namely, Les Houches angularity (LHA) [247], splitting variables r_{g} and z_{g} [248, 249], N-subjettiness ratios, τ_{21} and τ_{32} [250], and the ratios of energy correlation functions, C_2, D_2, ECF2, and ECF3.

LHA is the case where the exponents $\kappa = 1, \beta^{\mathrm{LHA}} = 0.5$ in the generalised angularity expression:

$$\lambda^{\kappa}_{\beta^{\mathrm{LHA}}} = \sum_{i \in J} z_i^{\kappa} \theta_i^{\beta^{\mathrm{LHA}}},$$

where z_i is the transverse momentum of jet constituent i as a fraction of the scalar sum of the p_{T} of all constituents and θ_i is the angle of the ith constituent relative to the jet axis, normalised by the jet radius.

Energy correlation functions ECF2 and ECF3 [251], and related ratios C_2, D_2 [252] for a jet J are derived from:

$$\begin{aligned}
\mathrm{ECF1} &= \sum_{i \in J} p_{\mathrm{T}_i}, \\
\mathrm{ECF2}(\beta^{\mathrm{ECF}}) &= \sum_{i<j \in J} p_{\mathrm{T}_i} p_{\mathrm{T}_j} \left(\Delta R_{ij}\right)^{\beta^{\mathrm{ECF}}}, \\
\mathrm{ECF3}(\beta^{\mathrm{ECF}}) &= \sum_{i<j<k \in J} p_{\mathrm{T}_i} p_{\mathrm{T}_j} p_{\mathrm{T}_k} \left(\Delta R_{ij} \Delta R_{ik} \Delta R_{jk}\right)^{\beta^{\mathrm{ECF}}},
\end{aligned}$$

where the parameter β^{ECF} weights the angular separation of the jet constituents. In this analysis, $\beta^{\mathrm{ECF}} = 1$ is used, and for brevity, β^{ECF} is not explicitly mentioned hereafter. The ratios of some of these quantities (written in an abbreviated form) are defined as :

$$e_2 = \frac{\mathrm{ECF2}}{(\mathrm{ECF1})^2},$$

$$e_3 = \frac{\mathrm{ECF3}}{(\mathrm{ECF1})^3}.$$

These ratios are then used to generate the variable C_2 [251], and its modified version D_2 [252, 253], which have been shown to be particularly useful in identifying two-body structures within jets [254]:

$$C_2 = \frac{e_3}{(e_2)^2},$$

$$D_2 = \frac{e_3}{(e_2)^3}.$$

The N-subjettiness describes to what degree the substructure of a given jet is compatible with being composed of N or fewer subjets. In order to calculate τ_{N}, first N subjet axes are defined within the jet by using the exclusive k_{t} algorithm, where the jet reconstruction continues until a desired number of jets are found. A parameter β^{NS} gives a weight to the angular separation of the jet constituents. In the studies presented here, the value of $\beta^{\mathrm{NS}} = 1$ is used.

Among these observables, this study primarily focused on C_2, LHA and τ_{21}, and τ_{32}İn general it was observed that D_2 and ECF2 are fairly similar, but are less sensitive as compared to C_2, and ECF3, r_g and z_g are mostly insensitive to the effect being probed.

In order to compare signal and background large-radius jets with similar kinematics, two different jet p_T ranges, 400-600 GeV and 800-1000 GeV are looked at, motivated by the leading jet p_T distribution in Fig. 5.3.

5.2.3 Results

Distributions of several jet substructure observables are compared between semi-visible and light quark or gluon initiated jets in Fig. 5.4. The results in leading jet p_T range of 400-600 GeV are shown, but the results in the 800–1000 GeV range exhibit the same feature, albeit with a lack of statistics. The distributions are normalised to area, not to cross-section, relevant for probing the shape differences.

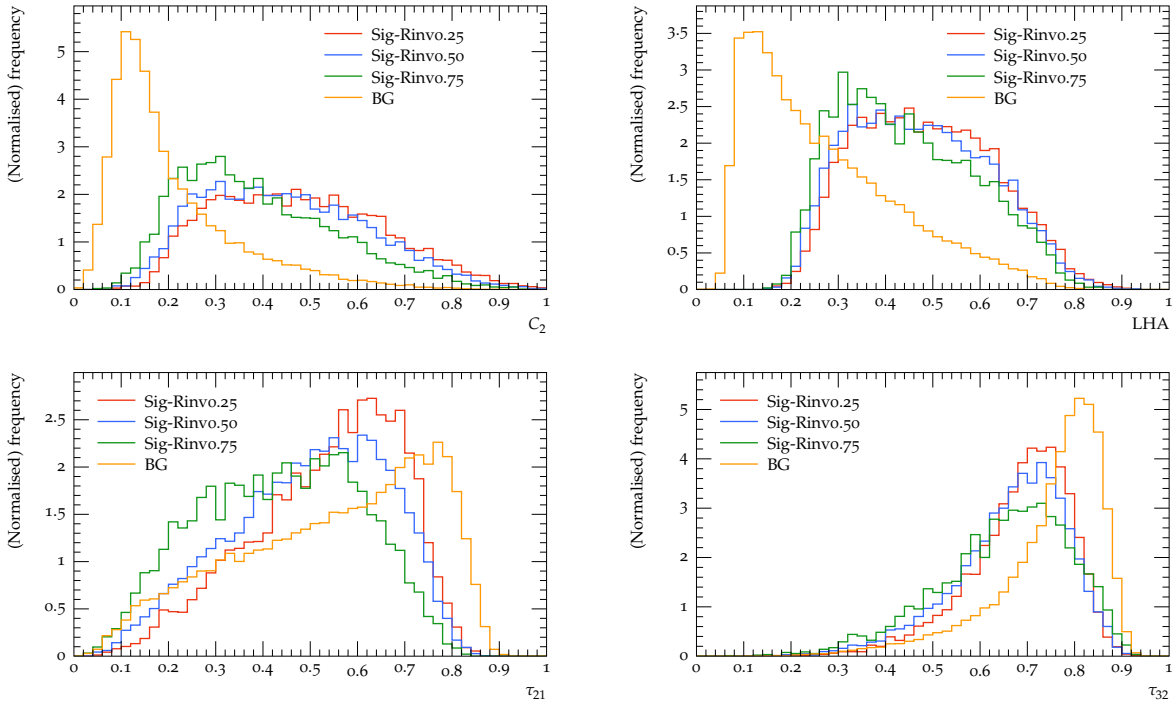


Figure 5.4: Comparisons of C_2 (top left), LHA (top right), τ_{21} (bottom left) and τ_{32} (bottom right) between three different signals corresponding to r_{inv} values of 0.25, 0.50 and 0.75, and the background for α_{FSR} of 1 and p_T between 400-600 GeV.

The overall interpretation is, SVJ result in more multi-pronged substructure, as evidenced in higher values of C_2 and LHA. For τ_{21}, and τ_{32}, the lower values of signal indicate that those are more multi-pronged respectively, whereas the background jets are more single-pronged. LHA, surprisingly does not show any difference when varying r_{inv} fractions. For N-subjettiness observables, the distributions with lower r_{inv} values have similar behaviour compared to background, indicative of the the fact that lower dark hadron fraction is less multi-prong, and indeed more background. It is important here to note that the N-subjettiness and ECF variables have somewhat different design philosophy [255]. While the former strongly depend on the axis, and are more sensitive to determine if the jet has at least N-prongs, the latter are more sensitive to separating one- or two-prong substructures.

Therefore, if the SVJ is more multi-pronged than two-pronged, then these two classes of observables can appear to show contradictory characteristics, i.e. τ_{21} value will indicate that the SVJ is atleast two-pronged, whereas C_2 will state it is not two-pronged. For example, C_2 for the signal with highest r_{inv} fraction appear closest to the background, in apparent contradiction with τ_{21}, which just implies that it is rather different from being two-pronged, not necessarily single-pronged.

The results here are shown without any theoretical systematic uncertainties. Based on the recent study [246], it is conservatively assumed that a 30–40% flat uncertainty can be applied on these substructure variables. That would not make the general conclusions arrived at this article invalid, but for certain observables, like τ_{21} for lower r_{inv} values, the discrimination power would be degraded. Also, detector effects can degrade the performance as well, but a quick check using parametrised smearing [256] showed the results obtained are robust. Smearing of the substructure variables makes the peaks somewhat diffused and the difference between the signal and background slightly less pronounced.

5.2.4 Understanding the model dependence

Currently the only dark shower model that can be used to simulate semi-visible jets is the Pythia 8 Hidden Valley module. So an obvious concern is, to what extent the differences seen between signal and background in the previous section is model-dependent. Due to the absence of another model, an unambiguous answer to this question is difficult to arrive at, but considering an extreme scenario of r_{inv} of zero might offer some hints. Imposing this condition implies that the signal large-radius jets consist entirely of visible hadrons, and subsequently the behaviour is expected to be like background jets, with low missing transverse momenta, as seen in Fig. 5.3.

However, in this case, requirements on missing transverse momentum magnitude and direction does not really make sense for signal, so for these comparisons, a background-like event selection is employed, assuming leading large-radius jet is the SVJ.

Considering a background-like event selection, along with the r_{inv} of zero condition, if the substructure of the signal jets resemble that of the background jets, then that would impart some confidence that the difference seen for non-zero r_{inv} values, as seen before, are caused not only by the model specifications but also involve the effects owing to the dark hadrons. The two parameters controlling the HV shower, that were expected to to be most consequential for this study, are defined in Table 5.1.

Observable	Pythia8 name	Indicated in text by:
Fixed alpha scale of gv/gammav emission	*HiddenValley:alphaFSR*	α_{HV}
Lowest allowed p_{T} of emission	*HiddenValley:pTMin*	$p_{\text{THV}}^{\text{min}}$

Table 5.1: Hidden Valley model parameters considered in the study. The HV fixed alpha scale corresponds to α_{strong} of QCD or α_{em} of QED.

Minimal dependence on $p_{\text{THV}}^{\text{min}}$ (which was also fairly independent of dark-hadron mass scale) is seen, but in Fig. 5.5, the substructure variables change significantly with the variation of α_{HV}, where other intermediate values were also probed, but are not shown.

The takeaway message is that in signal jets, C_2 can be made to look similar to background jets for α_{HV}= 0.1. The trend for LHA is not so clear, and the N-subjettiness observables have the weakest dependence, indicating the latter is not affected by the HV model implementations.

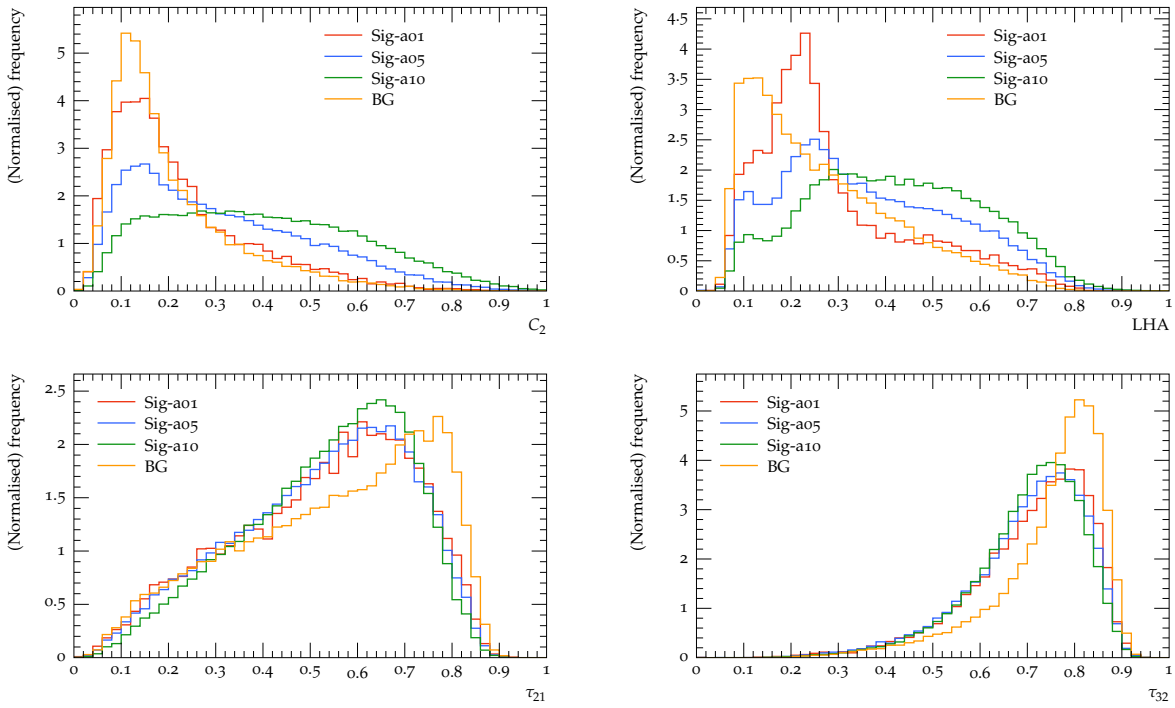


Figure 5.5: Comparisons of C_2 (top left), LHA (top right), τ_{21} (bottom left) and τ_{32} (bottom right) between three different signals corresponding to α_{HV} values of 0.1, 0.5 and 1.0 with r_{inv}= 0, leading jet p_{T} range 400-600 GeV, and the background. It is interesting to note that that a signal with r_{inv} of 0 is not necessarily equivalent to the background.

While this α_{HV} value is the closest to the QCD α_{FSR} value used in generators, it should be noted that they cannot be treated at the same footing, as QCD coupling is run at 2-loops. However, based on these results, this α_{HV} value is used in the rest of the comparisons.

5.2.5 Origin of the differences

For an understanding of the observed behaviour of jet substructure observables in SVJ, three questions come to mind:

1. What is effect of initial state radiation (ISR) and extra radiation on jet substructure?
2. Does decay from intermediate to final dark hadrons affect the substructure?
3. How does grooming affect jet substructure in SVJ?

In order to answer these, the other extreme scenario of $r_{\text{inv}}$of unity, which corresponds to the case where the signal jet consists entirely of dark hadrons is probed. Evidently in this case the signal jet itself is ill-defined, but by considering the unphysical scenarios of using dark hadrons in jet clustering, several effects can be disentangled.

First, the dark hadrons can be used to form signal jets, along with visible hadrons or without visible hadrons. The extra ME jets and the ISR can be turned off in either case. In each case, the leading large-radius jet is taken, and unless otherwise mentioned, comparisons are performed in the p_{T} range of 400–600 GeV. The same observables are studied as before in Fig. 5.6.

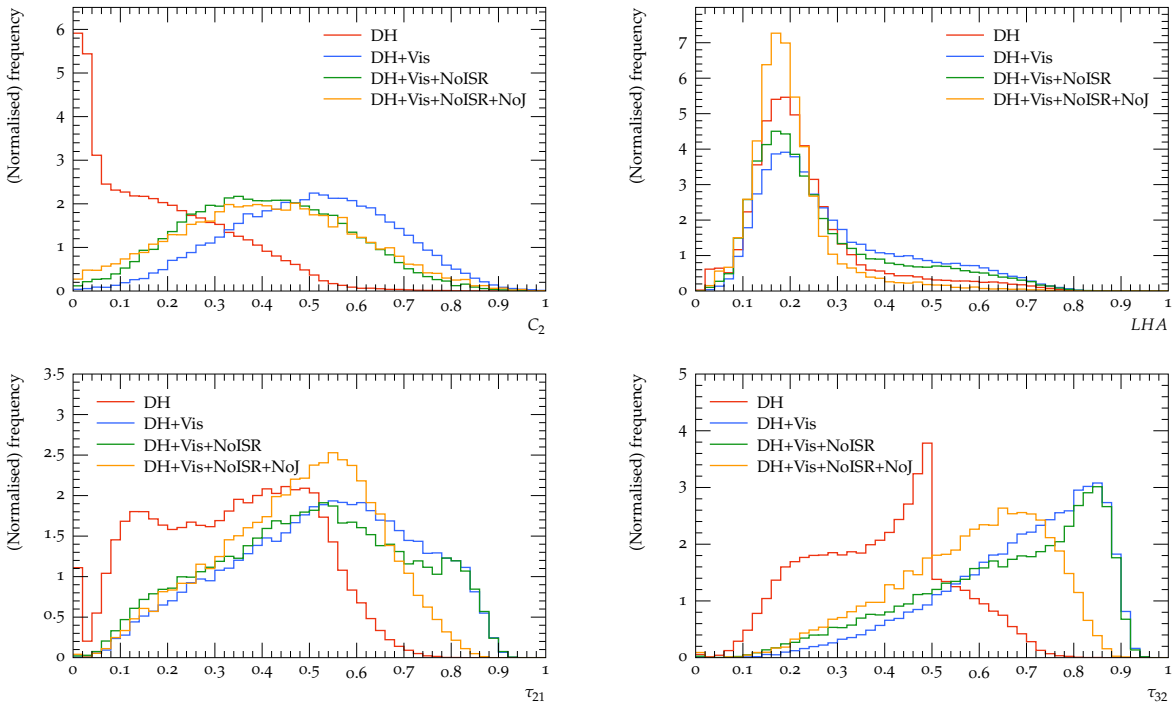


Figure 5.6: Comparisons of C_2 (top left), LHA (top right), τ_{21} (bottom left) and τ_{32} (bottom right) between different signals corresponding to clustering only with dark hadrons (DH), adding visible hadrons (Vis), tuning ISR off (NoISR), and also turning extra ME jets off (NoJ).

Clustering only dark hadron in jets is indicative of the shape an ideal SVJ may result in. The more realistic scenario is of course clustering the visible hadrons. In r_{inv} of unity scenario considered here, the visible hadrons come almost exclusively from ME level extra jets and ISR. Looking at C_2 and τ observables, its clear that adding visible hadrons make the signal jets less multi-prong, by filling in the gaps. As before higher C_2 values indicate moving away from two-prong structure.

It is interesting to see how the visible hadrons coming from ISR and ME extra jets affect the substructure differently. Turning off the ISR affects C_2 more than N-subjettiness observables, perhaps indicating the C_2 is more sensitive to the softer radiation. Additionally turning ME extra jets off has the opposite behaviour, it does not affect C_2, but makes N-subjettiness values indicative of slightly more two- or three-pronged substructure. It also implies that ISR adds more activity to semi-visible jets compared to ME extra jets, making them slightly more multi-pronged. This may be due to the fact ISR jets are more isotropic so they can overlap with SVJ, while ME jets are more well separated. Turning off ME extra jets makes the SVJ produced with less p_{T}, implying that the *same* jets are not compared in these cases. Surprisingly LHA seem rather insensitive.

An interesting feature can be seen the bottom left τ_{21} distribution of Fig. 5.6, where two peaks appear. This feature in enhanced for the higher p_{T} range, and also appears for lower values of α_{HV} as discussed in Sec. 5.2.4, which can be seen in Fig. 5.7. This is independent of adding SM hadrons, except when ME extra jets are turned off. This observation is consistent with the occurrence of this feature with higher p_{T}, where jets can be more collimated and two-pronged. The lower values of α_{HV} similarly indicate less radiation.

Another sanity check is to examine if the decay from intermediate dark hadrons to the

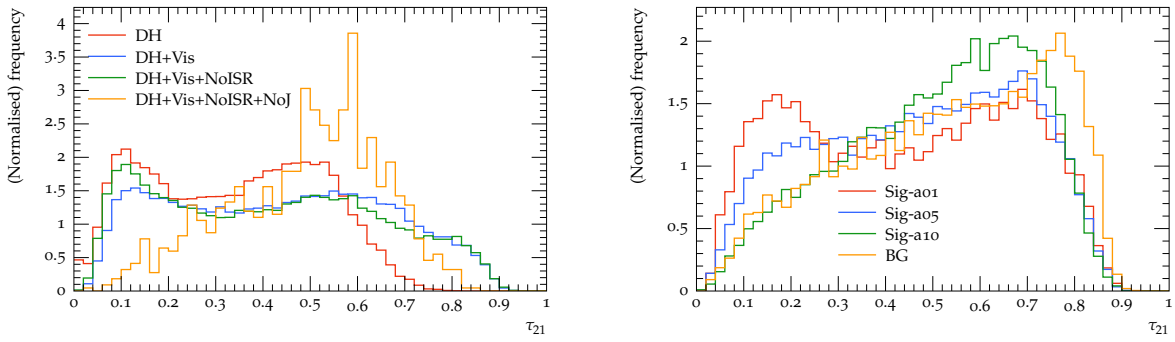


Figure 5.7: Comparisons of τ_{21} in leading jet p_{T} 800-1000 GeV range between different signals corresponding to clustering only with dark hadrons, adding visible hadrons, tuning ISR off, and also turning extra ME jets off with r_{inv} of 1 (left) and for three different signals corresponding to α_{HV}values of 0.1, 0.5 and 1.0 with $r_{\mathrm{inv}}$of 0 (right).

final dark hadrons considered above is responsible for creating or enhancing the substructure, by making the intermediate dark hadrons stable, and cluster them in jets, with and without visible hadrons. In Fig. 5.8, the comparison of those with the previous results show essentially no difference, except a slightly more flattish shape in lower values of N-subjettiness for the current case. So it is safe to say the observed substructure is not due to HV decay structure.

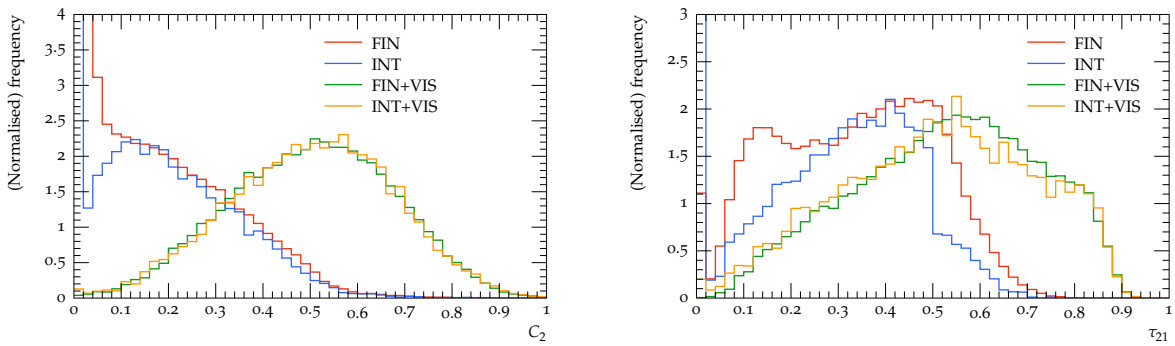


Figure 5.8: Comparisons of C_2 (left) and τ_{21} (right) between different signals corresponding to clustering only with final dark hadrons (FIN), intermediate dark hadrons (INT) and adding visible hadrons (VIS) in both cases. The entries at zero correspond to cases where the substructure variable could not be calculated, as only in rare cases the actual value of the observable was zero.

The next test was how grooming affects the substructure of SVJ, as grooming preferentially cuts out soft or wide angle radiation, and this study tested the effect of trimming.

In Fig. 5.9, different configurations with and without trimming are compared. Trimming in general moves τ_{21} values to the left, indicating a cleaner two-pronged substructure. This is least pronounced for 'only dark hadron' case, slightly more when visible hadrons are also clustered, and most pronounced for no extra ME jets or ISR case. A comparison between the scenarios of no extra ME jet and no ISR indicates ISR gets more affected by trimming. The same conclusion could also have been reached at looking at C_2, but the effect was less pronounced. Trimming did not affect the p_{T} spectra of the signal jets.

For finite r_{inv} values, only the visible hadrons are clustered in jets in Sec. 5.3.2, and slightly different substructure were seen for different r_{inv} values. Now, as in Sec. 5.2.5, if the final dark hadrons are also clustered in the jets, this difference is expected to disappear,

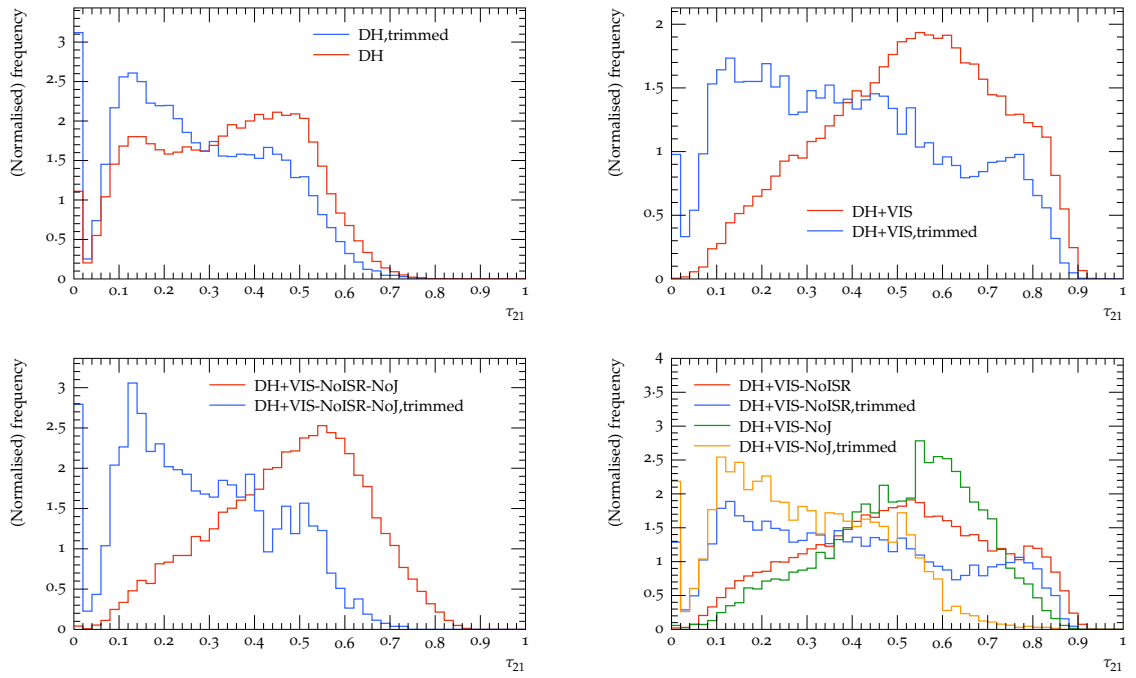


Figure 5.9: Comparisons of τ_{21} (right) between ungroomed and trimmed case, for signal configurations corresponding to clustering only with dark hadrons (top left), dark and visible hadrons (top right) turning extra ME jets and ISR off (bottom left) and turning one (but not both) of them off (bottom right). The entries at zero correspond to cases where the substructure variable could not be calculated.

as the different amount of *missing* hadrons in each case presumably was responsible for the difference, and that is the case as shown in Fig. 5.10. For C_2, the lines corresponding to the cases where dark hadrons are clustered are almost identical, and while they are not identical for τ_{21}, they lie in between the two original lines. This indicate that the substructure becomes less two-pronged with visible and dark hadrons in them, and the absence of the dark hadrons create the two-pronged structure. These distributions were made with α_{HV} of one to have the maximum possible dark radiation.

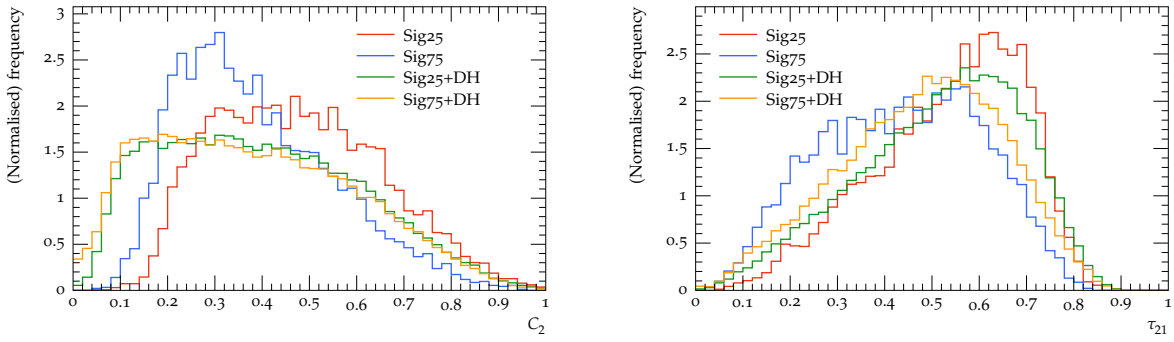


Figure 5.10: Comparisons of C_2 (left) and τ_{21} (right) for different signals corresponding to r_{inv} values of 0.25 and 0.75, clustering only the visible hadrons and clustering also with final dark hadrons.

5.3 Exploring new observables for dark sector

There always lies the possibility that the nature of unusual final states, like SVJ, are not well captured by the conventional JSS observables. This section discusses the exploratory study performed for checking the possibility of designing observable(s) to distinguish between semi-visible jets and light quark/gluon jets by comparing different observables. The families of variables feature angular scaling parameters that vary their sensitivity to different angular scales of jet emissions, potentially sensitive to the changes in jet structure introduced by dark-shower splitting, for various dark-hadron masses. Studying the variations of such observables, and their uncertainties between MC models and MC theory systematic uncertainties, will enable a comprehensive survey of how to maximise measurement sensitivity across the BSM model space; in particular, varying the DM mass might affect the shower, since it impacts the semi-visible splitting kinematics, and allow to design "theory-safe" variables directly motivated by the splitting structure. *This section is primarily based on Ref [257], unless stated otherwise.*

5.3.1 Energy flow polynomials

Energy flow polynomials [258] (EFPs) are observables that are multi-particle energy correlators with specific angular structures which directly result from IRC safety. EFPs form a linear basis of all IRC-safe observables, making them suitable for a wide variety of jet substructure contexts where linear methods are applicable. For a multigraph G with N vertices and edges $(k, l) \subset G$, the corresponding EFP takes the form

$$\mathrm{EFP_G} = \Sigma_{i_1=1}^{M} ... \Sigma_{i_N=1}^{M} z_{i_1} ... z_{i_N} \Pi_{(k,l)\subset G} \theta_{i_k i_l} \,, \tag{5.1}$$

where the jet consists of M particles, z_i is the energy fraction carried by particle i, and θ_{ij} is the angular distance between particles i and j.

Each edge (k, l) in a multigraph is in one-to-one correspondence with a term θ in an angular monomial. Each vertex j in the multigraph corresponds to a factor of z and summation over i_j in the EFP, as can be seen from Figure 5.11.

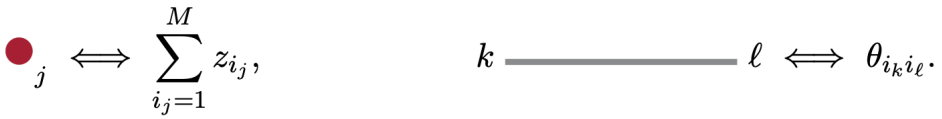


Figure 5.11: EFP construction: vertex and angular connectors

Hence, two particles/constituents of a jet can be treated as two energy fractions with a single angularity connection between them, leading to a degree-1 polynomial, as can be seen from Figure 5.12

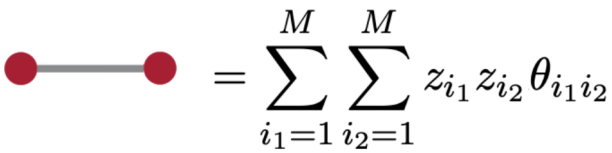


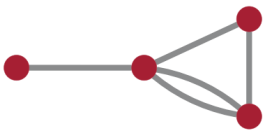
Figure 5.12: EFP construction: a degree one polynomial

$N=4$, $d=N-2$	0	1	2	3
0	-	-	-	-
1	1	-	1	2
2	-	-	-	-
3	-	-	1	-

Table 5.2: Grid formation, translation Figure 5.13 into a set of "particle" pairs.

Because the EFP basis is infinite, a suitable organization and truncation scheme is necessary to use the basis in practice. Several combinations of diagrams can be designed using the infinite number of vertex and particle correlator connections possible, however, the scope of this project has so far been restricted to exploring combinations with up to 7 particles/constituents/subjets and 8 angularity connectors between them.

EFPs are implemented in RIVET [239] framework and particular combination of EFPs are looked at to see if they help to distinguish between standard q/g jets and more unconventional jets. This in turn might lead to a new jet-substructure observable for dark shower discrimination. For this study, EFP multigraphs have been computed up to $N = 7$, $d = N-1, N, N+1$, which takes into account the different possible orientations of the input "particles" and designs an array of possible EFP diagrams as a grid. For the EFP diagram shown in Figure 5.13, a corresponding grid is designed as follows in Table 5.2, which translates the EFP to a set of "particle" pairs.



$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta^2_{i_2 i_4} \theta_{i_3 i_4}$$

Figure 5.13: EFP diagram with 4 constituents and 5 angularity connectors

Individual EFP diagrams use $R = 0.2$ anti-k_T subjets, from the leading jet in each event, as inputs and the log-likelihood ratio (LLR) value between pseudodata (Signal + SM background) and MC (SM background) is obtained after a very inclusive particle-level analysis with no $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold, jet $p_{\mathrm{T}} > 50\,\mathrm{GeV}$, $|\eta| < 4.9$, and vetoing leptons.

Certain EFP diagrams seem to have some bins of the jet-shape observables that QCD just doesn't populate at all, in which the DM signal dominates. On comparing several of these LLR distributions, as can be seen from Figure 5.14, some distinct LLRs have been identified that deviate from SM (here, multijet background is treated as a null-hypothesis), and the corresponding EFP diagrams are studied, as shown in Figure 5.15.

5.3.2 Results

It is necessary to understand what physics is being probed (i.e. the corresponding EFP equations) [259] and whether they are close to any standard jet substructure (JSS) variables. This is achieved by looking at correlations between the distinct EFPS and known JSS observables like N-subjettiness ratios [250], energy correlation functions (single ratios [251]), Les-Houches angularity (LHA) [247]. 2D Distributions of several JSS observables vs a selected EFP are compared between semi-visible and light quark or gluon initiated jets in Fig. 5.16 and 5.17. In particular, it is observed that there are distinct populations of signal

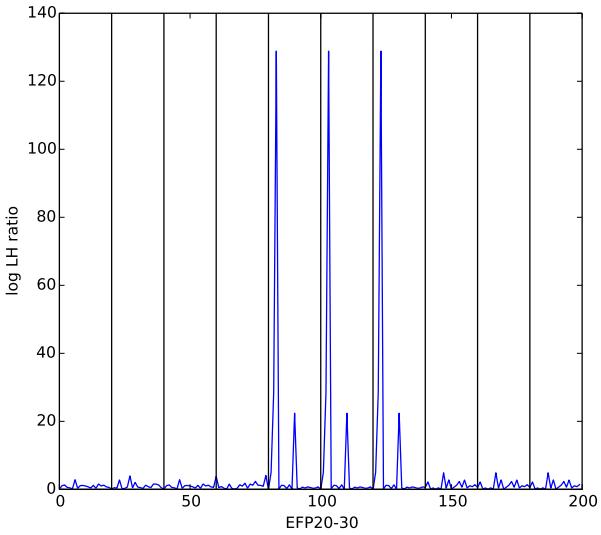


Figure 5.14: LLR summary distribution containing 10 EFP diagrams, showing a distinct pseudodata (Signal + SM background) and MC (SM background) difference in a few EFP diagrams (denoted by spikes).

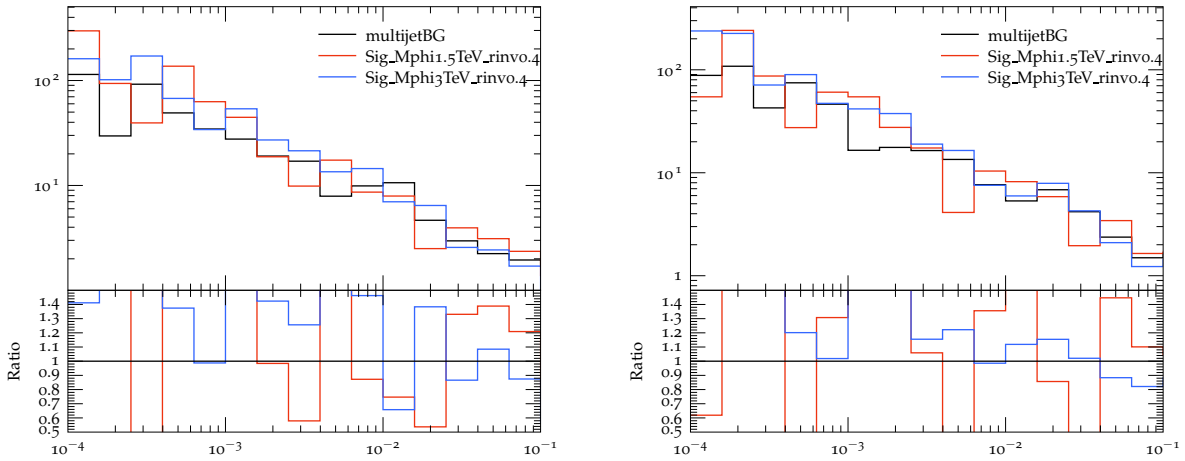


Figure 5.15: EFP distributions corresponding to spikes in LLR summary plot.

and background contributions at different ranges of C_2 and LHA. The very preliminary results show that the correlation of EFPs with a standard JSS observable has the potential to provide improved discriminating power between the signal and background.

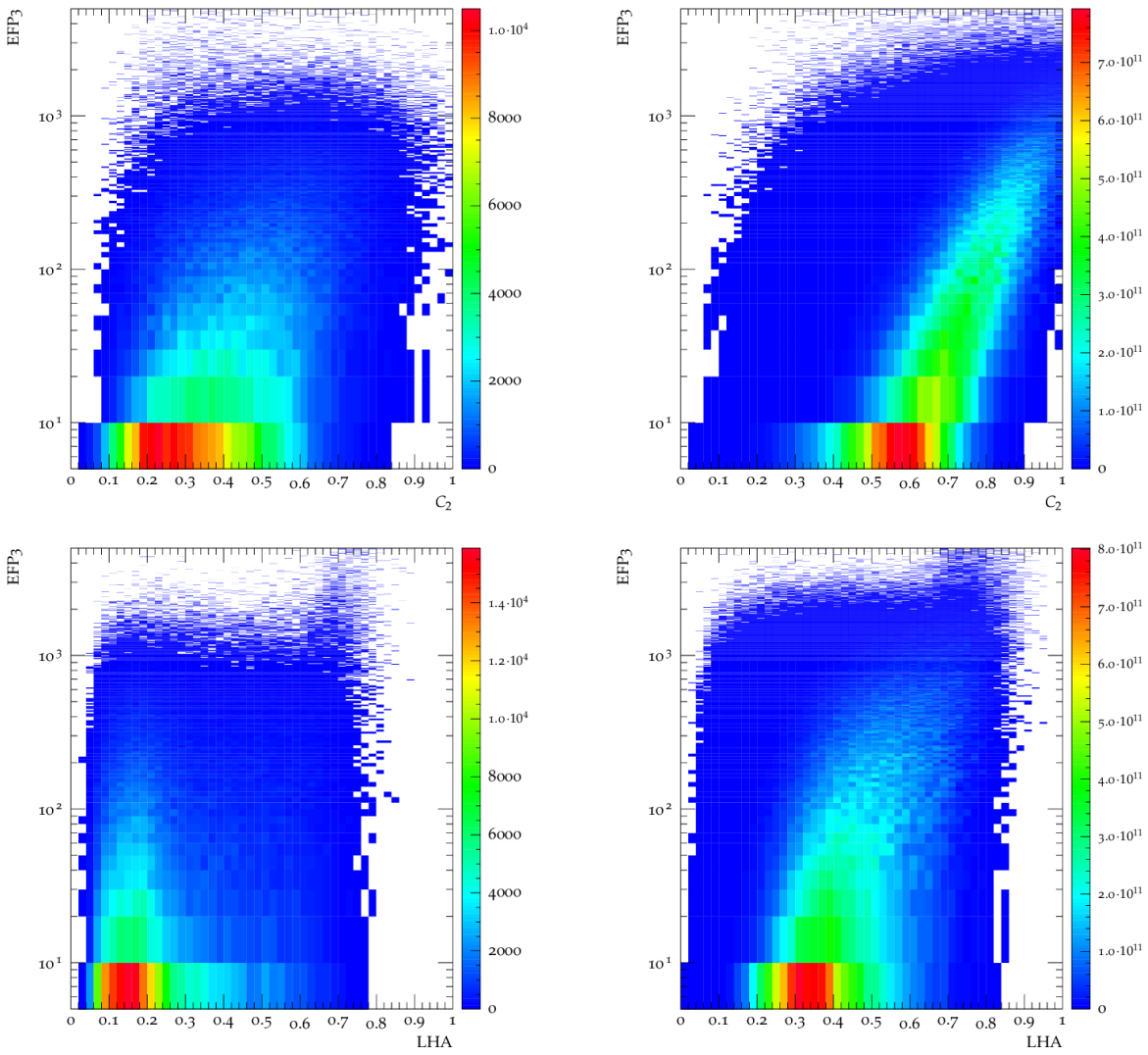


Figure 5.16: Comparisons of C_2 (top), LHA (bottom) with respect to EFP3 between a signal corresponding to r_{inv} of 0.3, and M_ϕ of 1.5 TeV (left) and the background (right).

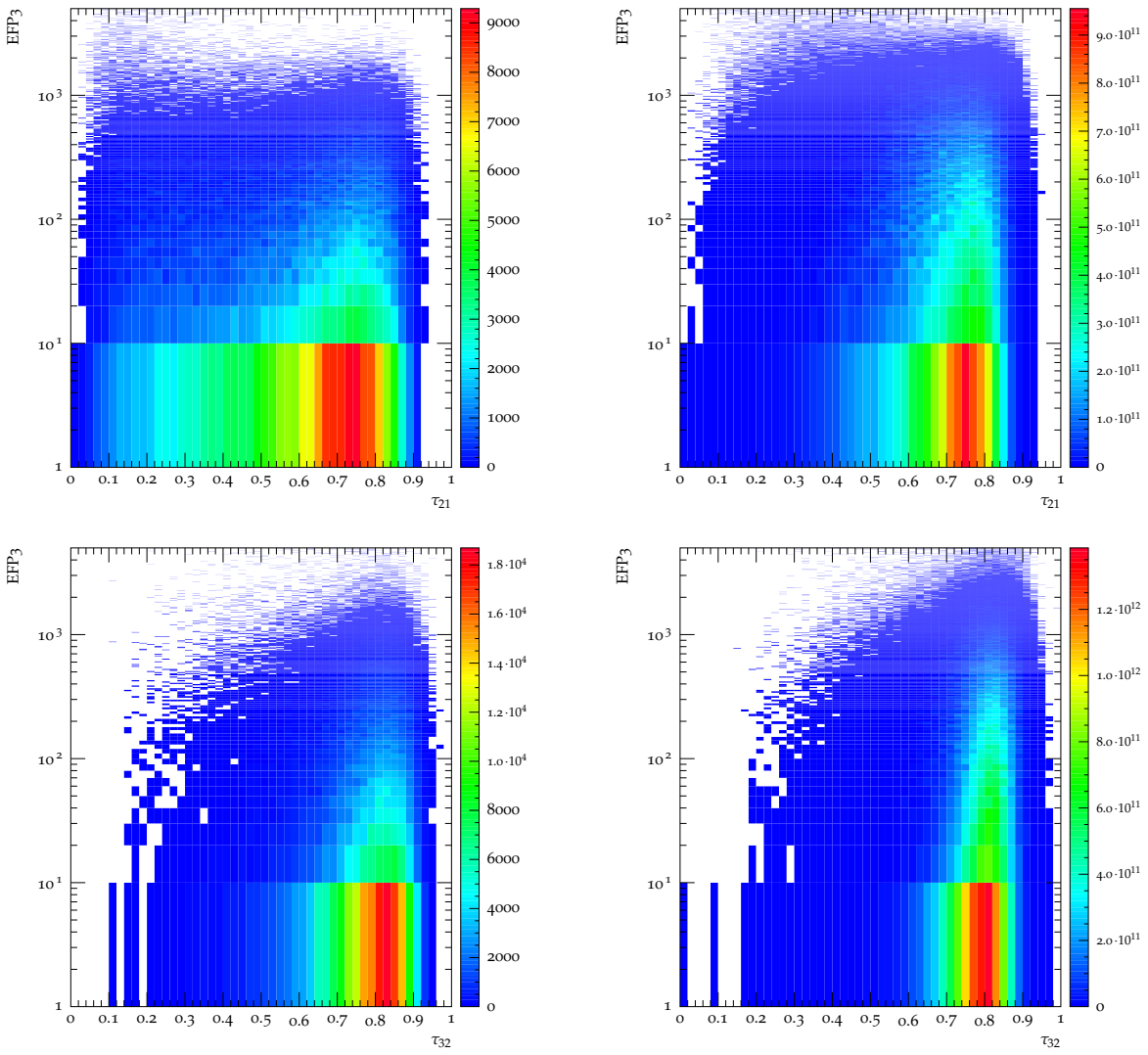


Figure 5.17: Comparisons of τ_{21} (top), τ_{32} (bottom) with respect to EFP3 between a signal corresponding to r_{inv} of 0.3, and M_ϕ of 1.5 TeV (left) and the background (right).

5.4 Semi-visible jet production with heavy flavour

In this section, a signature of SVJ only being produced with SM b-quarks in the t-channel is proposed. However, some of the conclusions derived here can also be applicable in a s-channel search. *This section is primarily based on Ref [260], unless stated otherwise.*

5.4.1 Signal modelling

The modelling of this final state signatures is performed using the HV module of Pythia 8, which was designed in order to study a sector which is decoupled from the Standard Model (SM). In HV module, the Standard Model gauge group G_{sm} is extended by a non-Abelian gauge group G_d, where the SM particles are neutral under G_d, but new HV light particles are charged under G_d and neutral under G_{sm}. The interactions between SM fields and the HV particles are allowed by TeV-scale operators. The simplest HV model [123, 261, 262] assumes the addition of a $U(1)$ x $SU(N_d)$ gauge group, with couplings g' and g_d, with the $U(1)$ being broken by a scalar $< \phi >$. Assuming the addition of two new quark flavours, and the presence of analogous neutral hidden valley pions π_d^0, the π_d^0 can decay to SM hadrons via the Yukawa coupling, κ. If π_d^0 are much lighter than the other hidden valley hadrons of the theory, a helicity flipping suppression forces the $\pi_d \to b\bar{b}$ to be dominant, provided the masses satisfy: $2m_b < m_{\pi_d^0} < 2m_t$. This is the same helicity flipping suppression as is observed in the case of SM π^+ decaying to $\mu^+\nu$ instead of $e^+\nu$. Hence, the b-philic mode of hidden valley pion decay will be preferred unless there is a hierarchical κ matrix that opposes the mass enhancement.

5.4.2 Signal reconstruction

The pre-selection used in recent ATLAS t-channel search [195] is considered as a starting point. Jets are constructed using the anti-k_t algorithm with a radius parameter of $R = 0.4$, using both charged and neutral inputs. The leading jet p_{T} is required to be greater that 250 GeV, while all the other jets are required to have p_{T} of at least 30 GeV. Events are required to have at least two jets, with the jet closest to the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction in azimuthal angle being termed the SVJ candidate, and it is required be within $\Delta\Phi < 2.0$ of the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction. It is observed that the the sub-leading jet is the SVJ candidate in most of the events. Events with charged leptons having p_{T} of 7 GeV or more are vetoed.

The pre-selection for the b-tagged SVJ case follows the ATLAS analysis pre-selection. However, the 200 GeV $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement severely reduces the signal statistics, so in order to investigate the strategies to increase the signal efficiency, the $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement has been removed in this study. At detector level, mis-measurement of jets typically increase the $E_{\mathrm{T}}^{\mathrm{miss}}$, so this is not a completely unrealistic assumption. However, for the actual analysis, the 200 GeV $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement will be imposed, assuming that the conclusions derived here hold. The lepton veto with the lowest possible p_{T} is appropriate for the analysis dominated by light-quark initiated jets, however semi-leptonic decays from b-quarks indicate the events will have leptons, so that selection has to be optimised as well.

In order to arrive at an analysis strategy which can maximise the signal efficiency for this specific final state, it is worthwhile to investigate different jet reconstruction strategies. This final state offers a clean playground, as the SVJs need to b-tagged, unlike for the democratic production of all five flavours where the only handle is the azimuthal separation from the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction, which can introduce ambiguities.

- The ATLAS analysis uses jets with radius parameter of 0.4. The use of $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger

allows the use of jets with p_{T} of 30 GeV, which would not be possible with lowest un-prescaled single jet triggers, which currently in ATLAS requires a leading jet p_{T} of 450 GeV [263].

- The jet energy calibration for large-radius jets with anti-k_t algorithm using a radius parameter of 1.0 with trimming [243] as used in ATLAS [264] typically require a minimum p_{T} of 200 GeV. Two b-tagged large-radius jets with p_{T} of at least 200 GeV are required. The b-tagging of large-radius jets in ATLAS are typically performed with associated track jets [265].

- Another possible option will be to use reclustered jets (RC) [200], as that allows one to use lower p_{T} jets than usual large-radius jets, but in the cases RC jets consist only one small-radius input jet, the experimental jet mass calibration tends to be ill-defined. It is observed that this is indeed the case for a significant number of events for radii of 0.8 to 1.5. Since the radius is fixed, it is difficult to avoid this problem.

- Jet reconstruction with a variable radius (VR) [266] was introduced to increase the signal reconstruction efficiency for boosted resonance searches. The VR algorithm needs the minimum and maximum allowed jet radius, as well ρ, which is the mass parameter, resulting in the effective radius of the VR jets to scale as $R\ \rho/p_{\mathrm{T}}$. While different choices can be made for input jets to VR algorithm, starting from trackjets (jets only with charged particles at particle level) where going to much lower p_{T} is feasible, anti-k_t jets are used with radius parameter of 0.4 with a minimum p_{T} of 30 GeV as input to VR algorithm. The trackjets do not capture the totality of the SVJ, which leads to non-optimal performance for signal to background discrimination later. The minimum radius was kept at 0.4 and the maximum was set to 1.5 to stay well within the central part of the detector. The suggested value of ρ is $< 2p_{\mathrm{T}}$ in case of resonance searches, here 200 GeV was used. It was checked that using a lower value did not affect the shape of kinematic distributions, but reduced the acceptance. This can be understood from the fact that SVJ indeed behaves like a multi-prong jet [242]. Since VR jet radius adaptive compared to RC, VR jets are preferred over RC jets. ATLAS has used VR tracks jets in previous searches [267].

In order to compare the performance of the above mentioned jet reconstruction strategies, a metric has to be decided upon. As the pre-selection will require a b-tagged SVJ close to $E_{\mathrm{T}}^{\mathrm{miss}}$ direction, after requiring at least two b-tagged SVJ, it is required at least one of them must be within $\Delta\Phi < 2.0$ of the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction, without demanding any higher leading jet p_{T} threshold. Then in Figure 5.18, the multiplicity of b-tagged jets for jets with radii 0.4 and 1.0, as well as of VR jets is presented.

The general trend is, use of large-radius jets decrease the signal efficiency, however, using VR jets, even without any optimisation results in an improvement, at least after b-jet multiplicity requirement. In order to understand the above mentioned signal efficiency gain with VR, the objects in $\eta-\phi$ plane for four representative events are looked at, in Figure 5.19. It is evident VR better reconstructs the SVJs better than jets with radius of 0.4, both in terms of containing all the decay products, as well as being closer to the direction of missing transverse momentum. This is seen to be true for both the signal points with rather different r_{inv} fraction.

The next step is to optimise the pre-selections with VR jets to obtain further improvement in signal efficiency, as well as to obtain a good signal to background discrimination.

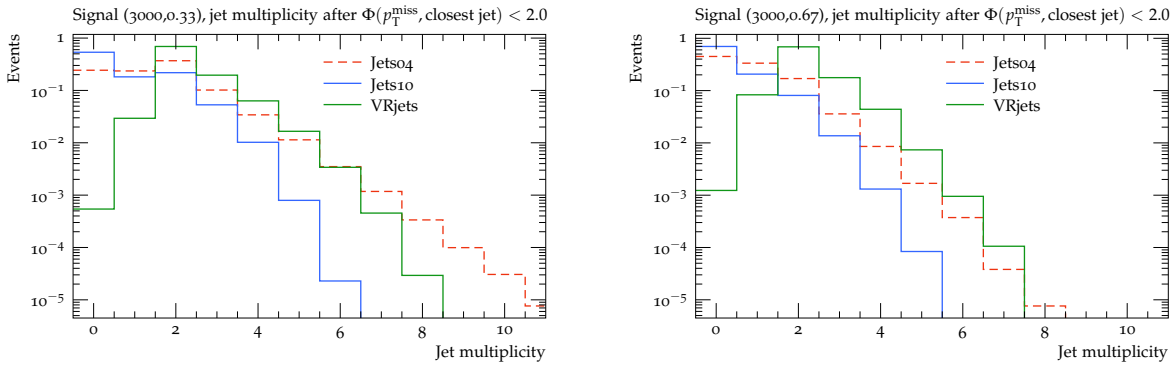


Figure 5.18: The multiplicity of b-tagged jets for radius 0.4 and radius 1.0 jets, as well as VR jets after $\Delta\Phi < 2.0$ requirement, for signal with r_{inv} of 0.33 (left) and 0.67 (right).

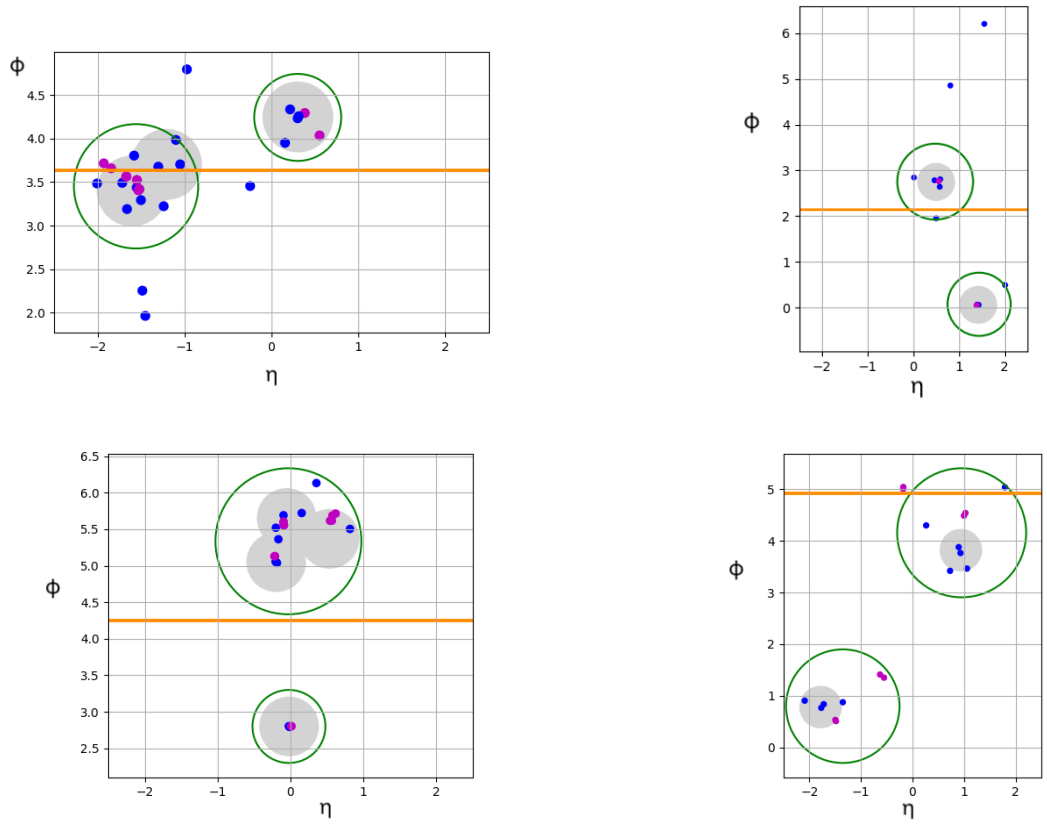


Figure 5.19: Various objects are plotted in $\eta - \phi$ plane for four representative signal events with r_{inv} of 0.33 (top row) and 0.67 (bottom row). The large hollow green circles represent the VR jets, the filled gray circles represent anti-k_t jets with R=0.4, the magenta points represent dark hadrons, the blue points indicate stable b-hadrond, and the orange line the direction of missing transverses momentum.

5.4.3 Search strategy

The pre-selections include discarding events with charged leptons having p_T of 7 GeV or more. While that is reasonable for a mostly light-flavour quark dominated signature, semi-leptonic decay of b-quarks produce a copious amount of charged leptons. Even in our particle level study, this requirements leads to a loss of about 50% of signal events with the 7 GeV

p_T threshold, dropping to about 25% for a higher p_T threshold of 30 GeV, which is highly non-ideal. To recover events from lepton veto, two approaches can be adopted, either vetoing events with charged leptons with a significantly higher p_T threshold, or requiring them to be close to the b-jet. In Figure 5.20, the correlation between these two quantities are shown. It can be seen most leptons have low p_T and are close to a b-tagged jet. Based on this, leptons were selected only with $p_T > 75$ GeV, and further require they lie within $\Delta\Phi < 0.1$ of a b-tagged jet. This leads to no signal efficiency loss, but still rejects a large fraction of background events with an isolated high p_T lepton.

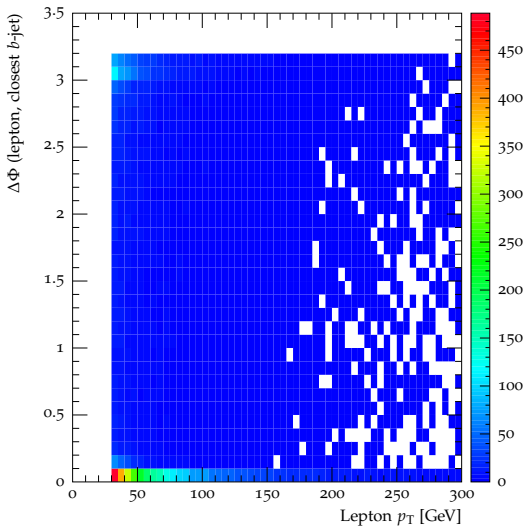


Figure 5.20: The correlation of charged lepton p_T against the $\Delta\Phi$ distance from the closest b-tagged jet is shown for the signal a mediator mass of 3000 GeV and r_{inv} of 0.33. Same trend is observed for the signal with r_{inv} of 0.67.

The leading background processes are multijet, and top quark pair production, referred to as $t\bar{t}$. The semileptonic and dileptonic decay modes of $t\bar{t}$ were seen to produce almost identical kinematic distributions and yields, and hence were combined together, termed as non-hadronic contributions. The jet closest in azimuthal angle to the direction of the missing transverse momentum is termed the SVJ candidate. The discriminating variables are similar to the ones listed in the ATLAS analysis. Previously it was found that the $p_T^{bal}$and $|\phi_{max} - \phi_{min}|$ are sufficiently different between signal and most background processes [195], so at the first step, the other observables in Figure 5.21 are checked.

Based on these distributions, it is required that the leading jet p_T be at least 80 GeV, with $\Delta\phi$(closest jet, E_T^{miss}) of 0.5 or smaller. Finally, the signal region (SR) can be defined with H_T of more than 400 GeV. in Figure 5.21. After these requirements, the distributions of $|\phi_{max} - \phi_{min}|$ and p_T^{bal} are shown in Figure 5.22. Unfortunately, it appears the former is not very helpful in this case, but the latter is a good discriminating variable, and can be used in the search.

For completeness, a cutflow has been provided in Table 5.3 summarising the above steps. The simple selections applied are already rather effective in defining a signal-enriched region, with p_T^{bal} being a sensitive observable. A detector level search will also have a requirement on E_T^{miss}, typically at least 200 GeV to reach the trigger efficiency plateau, which might further help in enhancing the signal over the background.

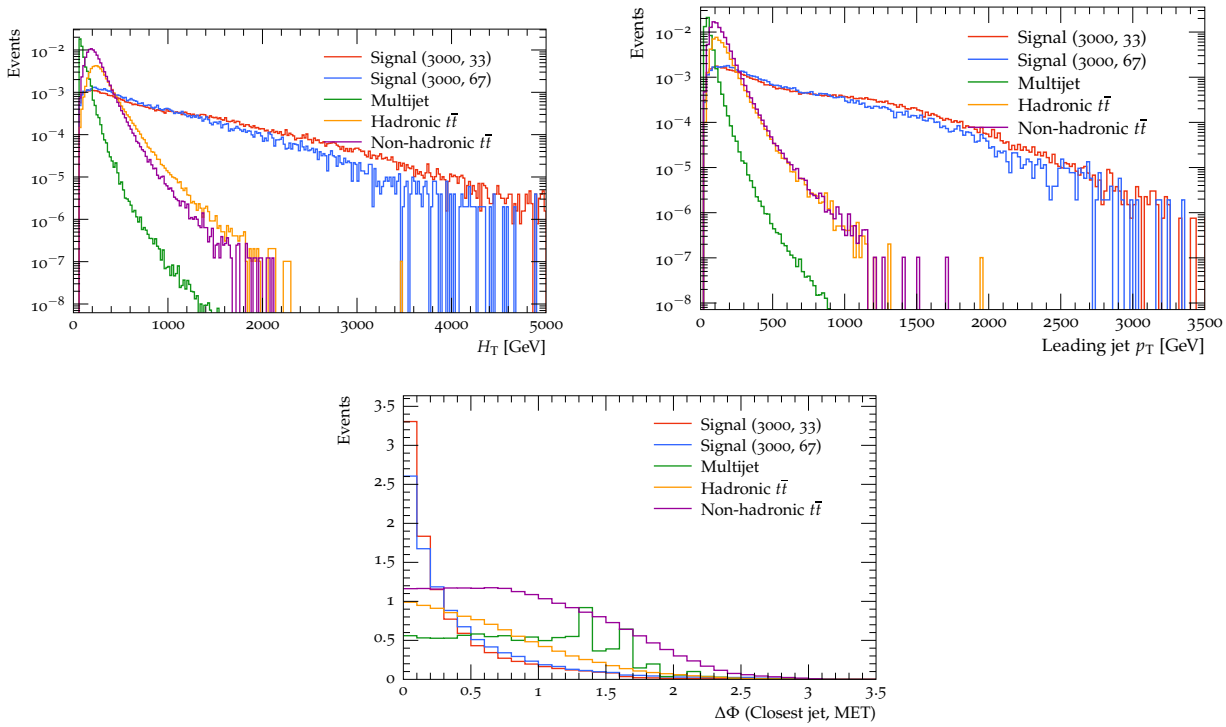


Figure 5.21: The distributions of H_T (top left), leading jet p_T (top right) and $\Delta\phi$(closest jet, E_T^{miss}) (bottom) for two signal (corresponding to a mediator mass of 3000 GeV and r_{inv} of 0.33 and 0.67) and all leading background processes after only jet multiplicity and lepton angle requirements.

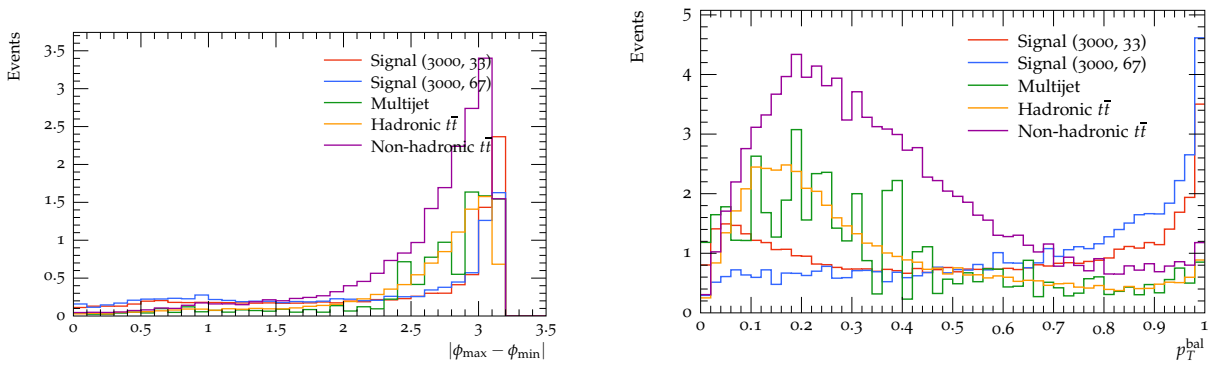


Figure 5.22: The distributions of $|\phi_{max} - \phi_{min}|$ (left) and p_T^{bal}(right) for two signal (corresponding to a mediator mass of 3000 GeV and r_{inv} of 0.33 and 0.67) and all leading background processes after all requirements in the SR defined by $H_T > 400$ GeV.

Requirement	Selection efficiency reduction in %:					
	Signal (3000,33)	Signal (3000, 67)	Multijet	Hadronic $t\bar{t}$	Semileptonic $t\bar{t}$	Dilpetonic $t\bar{t}$
DiBjet	51	20	1	49	49	47
Lepton angle	49	19	1	49	41	34
$\Delta\phi$(closest jet, E_T^{miss})	37	13	0	22	13	9.
Leading jet p_T	36	13	0	20	11	7
H_T SR	25	8	0	4	1	1

Table 5.3: The numbers represent the percentage of events remaining after each requirement. The individual requirements are mentioned in the text. The signal corresponds to a mediator mass of 3000 GeV and r_{inv} of 0.33 and 0.67, as indicated in the parenthesis.

5.5 Conclusions

There are several avenues of strongly interacting dark sector that can be explored, specifically looking into unusual final state signatures. This chapter discussed three different approaches to enhancing the sensitivity of final states containing SVJs in prospective collider searches.

Firstly, a comprehensive study of the substructure of semi-visible jets has been performed. It was demonstrated that specific hidden valley parameter configurations can reduce the dark shower model dependent features of the signal jets. The origin of the substructure in SVJ is neither caused by the decay of intermediate dark hadrons, nor by extra ME jets, or ISR, although the latter two affect the substructure. The substructure is created by the interspersing of visible hadrons with dark hadrons. The substructure observables which are least affected by model dependence can be used in searches, and also as inputs to machine learning algorithms trying to identify semi-visible jet via anomaly detection [258, 268–274], assuming the relatively similar contribution from signal and background processes.

The possibility of probing unique phase-space corners by exploiting the wealth of existing and new JSS observables in a IRC-safe linear basis, ala EFPs have been proposed. Standalone EFPs or combination of several EFPs can be correlated to an existing JSS observable and lead to improved discriminating power between the signal and background. Next steps in the study involve providing concrete recommendations for the combinations that can be utlised in explorations of the dark-sector at the LHC.

Finally, a feasibility study for a collider search for SVJ produced in association with only heavy flavour quarks have been presented. While this is theoretically well motivated scenario, no search has been performed yet, but it is a promising search channel. The additional requirement of SVJs being b-tagged acts as a powerful tool to isolate signal-like events, while the use of VR jets improves signal acceptance.

Chapter 6

Monte Carlo Truth Classifier

This chapter discusses the ATLAS authorship qualification task of implementing a tool widely used by the collaboration to study particle level objects for performance studies and physics analyses. It should be noted that throughout this chapter, the term "truth" will be used, which is the informal name used to refer to the particle level event record written out by Monte Carlo event generators.

6.1 Necessity of proper truth definitions

The LHC aims to provide precision measurements of theory predictions, and that requires an accurate definition of the observables used in physics measurements. These fiducial[1] definitions should enable the comparison between theoretical predictions and experimental results, while remaining as independent from the Monte Carlo event generators as possible. Such an observable definition also allows future theoretical developments to be unambiguously compared to current ones, with minimal prior knowledge of experimental or model-dependent definitions of the final state objects. Almost all measurements are ultimately made from the stable final state particles, however in some cases it is acceptable to use also their parents, as will be discussed later on.

Following the Truth Particle Workshop in 2014, recommendations were drafted [275] summarising the main discussion points of the workshop about the use of the truth objects in performance studies and physics analyses. Ref. [275] proposes a set of definitions for the most commonly used generator-level particles, based on information available in most Monte Carlo event generator records, and also shows examples of observables derived from these definitions.

In ATLAS, combined performance (CP) groups provide various corrections for reconstructed objects such as efficiency scale factors, MC calibrations or *in situ* calibrations. When computing scale factors, the matching of reconstructed objects to truth objects is used by the CP groups to avoid performing a "background subtraction" on MC samples. In such cases, it is recommended that CP groups choose a well defined truth object as discussed above. Especially for precision analyses it is beneficial if the same truth object definition is used for both the calibration and the unfolding[2].

[1] This implies the stable particles which account for the majority of interactions with the detector volume.

[2] Unfolding refers to a procedure of correcting the data for detector effects to particle level distributions, using MC simulations.

6.2 Original Monte Carlo Truth Classifier and drawbacks

Monte Carlo Truth Classifier (MCTC) is a tool used within ATLAS, to classify the truth particles according with their origin. Event generation step in ATLAS produces a format called EVNT, which contains the truth record in HepMC [276] format, wrapped in a form readable by ATLAS software, Athena. In detector-level events the truth information is retained, and this format is called xAOD. MCTC can be run on both EVNT and xAOD formats.

Based on the truth particle classification, the tool provides classification of inner detector and combined muon tracks, electrons, photons and jets. To classify truth particles, the tool defines their origin and provide simple classification based on this. The following attributes are assigned to each truth particle after classification:

- the particle type
- the particle origin
- the particle outcome process (or the particle final state).

The Particle Data Group's "Monte Carlo Numbering Scheme" identifier [277] is used for the definitions. Hadrons have their type classified as B-C-S-Light mesons or baryons but no origin or outcome is provided, whereas neutrinos and nuclear fragments are not classified. The particles with barcode[3] $> 10^6$ are classified as none primary, i.e. arising from detector material. All other particles are classified as generator particles. The electrons, muons, taus and photons are further subdivided into three main categories:

- Isolated (or prompt) - it is a primary particle, i.e., electrons from $Z \rightarrow ee$ decay,
- None Isolated - which originated from decay of the charm or bottom hadrons, and as an exception decay of J/ψ is also included
- Background - all the rest including bb and cc meson decays

If the classification algorithm has failed or there is no original vertex, the particle is classified as *Unknown* (electron, muon, tau or photon). Fig 6.1 shows the visual overview of the classification scheme employed by MCTC.

The original implementation however had a few drawbacks:

- There is no classification between prompt and non-prompt neutrinos as mentioned above.
- The tool has grown organically over the years and has diverted from the truth recommendations provided internally in the collaboration.
- The tool has been patched over the years for generator specific requirements
- The definition of "IsolatedElectron" has no relation to the normal definition of the term. It is rather a proxy for promptness of the particle.

[3]The ID assigned to truth particles in ATLAS, denoting detector interactions.

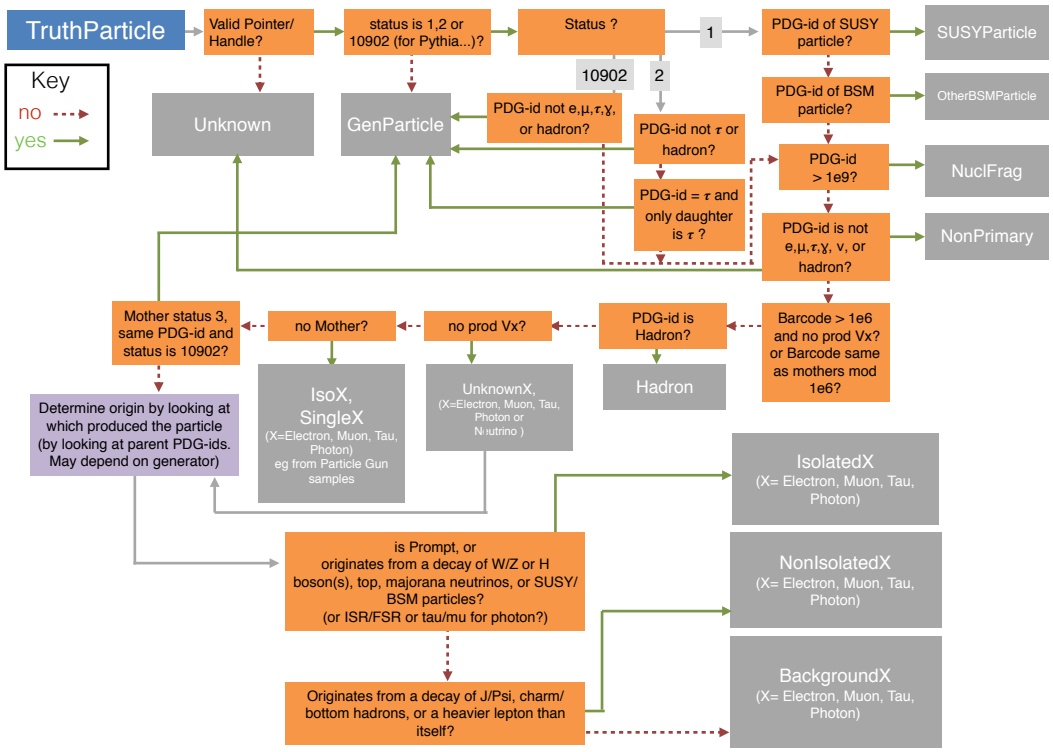


Figure 6.1: Flowchart of the old MCTC classification scheme.

- Final state radiation against initial state radiation classification is not recommended since the two effects can interfere, and the distinction is not always possible due to interference effects. The original implementation of MCTC has this unphysical distinction.
- Counting the number of quarks and gluons to determine if a particle came from e.g. a multi-boson event, is a highly problematic way of approaching the classification.

6.3 New implementation of MCTC

A new function (termed "Classify") has been defined which checks for hadron, τ, BSM, Geant4, or uncategorised origin and returns the output as a bitset converted to unsigned integer. The following boolean flags are used:

- isStable: Whether the particle is stable (i.e. status 1) or not [1/0].
- isGeant: the particle originates in a nuclear interaction or not [1/0].
- isBSM: the particle itself is a BSM particle or not [1/0].
- uncat: HepMC production vertex not found / found [1/0].
- fromBSM: a BSM particle in the ancestor chain or not [1/0].
- fromhad: a hadron in the ancestor chain or not [1/0].
- fromTau: a τ lepton in the ancestor chain or not [1/0].

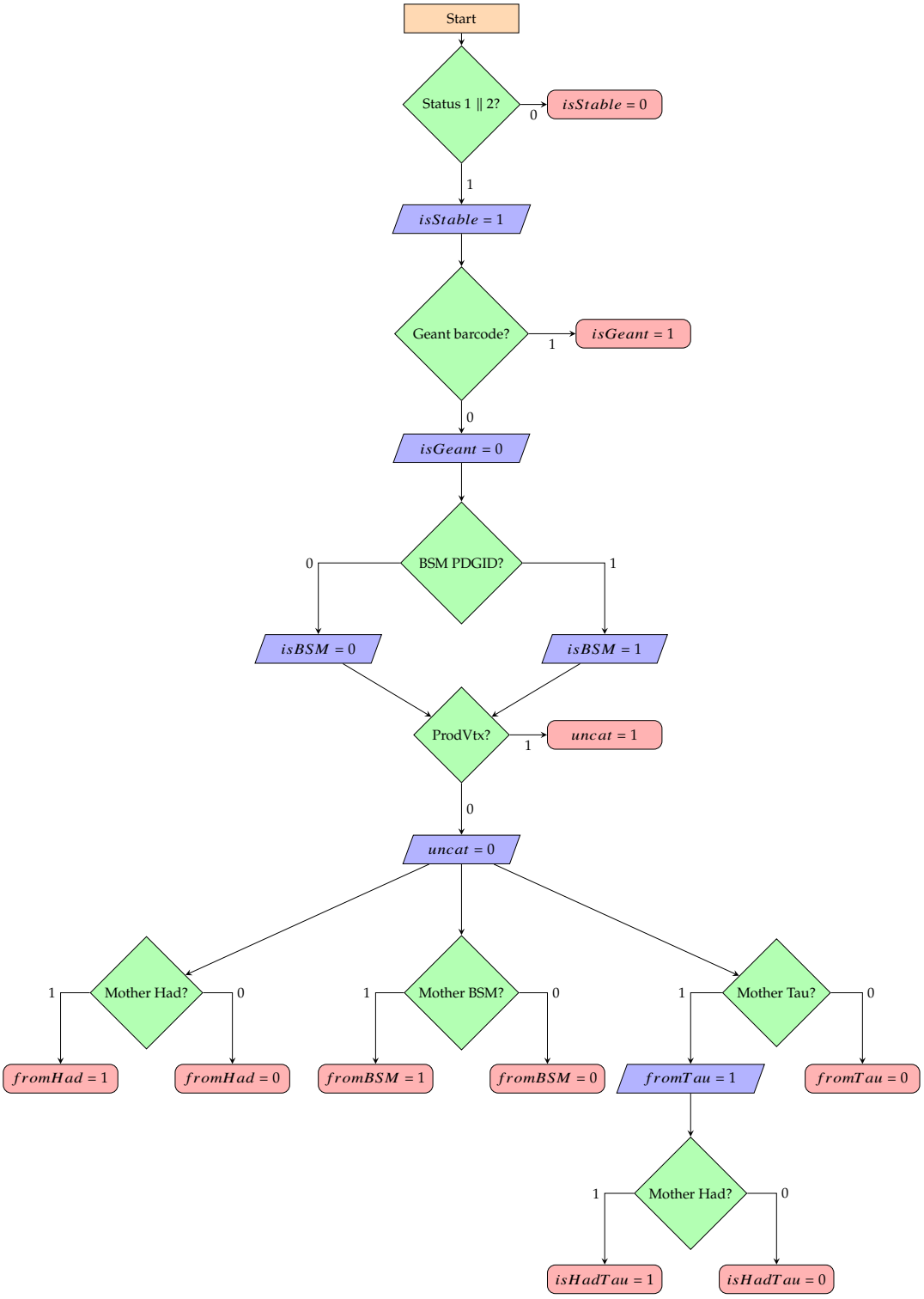


Figure 6.2: Flowchart of the new MCTC classification scheme.

- isHadTau: a τ lepton with a hadronic mother in the ancestor chain or not [1/0].

The logic flow is highlighted in Fig 6.2. The function uses boolean to check Prompt or non- prompt τ (isHadTau), Prompt or non-prompt (fromhad), whether it's coming from τ (fromTau), whether it is a BSM particle or from a BSM particle (isBSM and fromBSM), whether it is a particle coming from detector material interactions (isGeant) and to walk back the decay chain till beam particle, i.e proton is retrieved (frombeam). The resulting bitset has bits in the following order:

isStable	isGeant	isBSM	uncat	fromBSM	fromhad	fromTau	isHadTau

Several validation tests were performed, however only one is highlighted here as an illustration. In this example, events with a Higgs boson produced in association with top quark pair is selected. This physics process was chosen since it encompasses most of the use cases covered by the tool, depending on whether the Higgs boson or top quarks decay hadronically or semi-leptonically. An individual event when passed through MCTC returned a bitset value of 110100. When translated using the scheme mentioned above, this results in the following decay chain: b-meson $\rightarrow \tau \rightarrow e$, which is in line with the decay chain expected from this process. This implementation of MCTC by the thesis author is currently the recommended version of the tool that is being used in all ATLAS analysis frameworks.

Bibliography

1. Weinberg, S. The Making of the standard model. *Eur. Phys. J. C* **34** (ed 't Hooft, G.) 5–13. arXiv: hep-ph/0401010 (2004).
2. Glashow, S. L. Partial-symmetries of weak interactions. *Nuclear Physics* **22,** 579–588. ISSN: 0029-5582. https://www.sciencedirect.com/science/article/pii/0029558261904692 (1961).
3. Higgs, P. W. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.* **13** (ed Taylor, J. C.) 508–509 (1964).
4. Englert, F. & Brout, R. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.* **13** (ed Taylor, J. C.) 321–323 (1964).
5. Guralnik, G. S., Hagen, C. R. & Kibble, T. W. B. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.* **13** (ed Taylor, J. C.) 585–587 (1964).
6. Weinberg, S. A Model of Leptons. *Phys. Rev. Lett.* **19,** 1264–1266 (1967).
7. Salam, A. Weak and Electromagnetic Interactions. *Conf. Proc. C* **680519,** 367–377 (1968).
8. Wilson, K. G. Confinement of Quarks. *Phys. Rev. D* **10** (ed Taylor, J. C.) 2445–2459 (1974).
9. Zyla, P. A. *et al.* Review of Particle Physics. *PTEP* **2020,** 083C01 (2020).
10. Weinberg, S. New approach to the renormalization group. *Phys. Rev. D* **8,** 3497–3509 (1973).
11. 't Hooft, G. & Veltman, M. J. G. Regularization and Renormalization of Gauge Fields. *Nucl. Phys. B* **44,** 189–213 (1972).
12. Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. in *The standard theory of particle physics: Essays to celebrate CERN's 60th anniversary* (eds Maiani, L. & Rolandi, L.) 255–274 (2016). arXiv: 1504.07217 [hep-ph].
13. Peskin, M. E. & Schroeder, D. V. *An Introduction to quantum field theory* ISBN: 978-0-201-50397-5 (Addison-Wesley, Reading, USA, 1995).
14. Schwartz, M. D. *Quantum Field Theory and the Standard Model* ISBN: 978-1-107-03473-0, 978-1-107-03473-0 (Cambridge University Press, Mar. 2014).
15. Campbell, J., Huston, J. & Krauss, F. *The Black Book of Quantum Chromodynamics: A Primer for the LHC Era* ISBN: 978-0-19-965274-7 (Oxford University Press, Dec. 2017).
16. Field, R. D. *Applications of Perturbative QCD* (1989).
17. Collins, J. C. & Soper, D. E. The Theorems of Perturbative QCD. *Ann. Rev. Nucl. Part. Sci.* **37,** 383–409 (1987).

18. Soper, D. E. Parton distribution functions. *Nucl. Phys. B Proc. Suppl.* **53** (eds Bernard, C., Golterman, M., Ogilvie, M. & Potvin, J.) 69–80. arXiv: hep-lat/9609018 (1997).

19. Salam, G. P. Towards Jetography. *Eur. Phys. J. C* **67,** 637–686. arXiv: 0906.1833 [hep-ph] (2010).

20. Deur, A., Brodsky, S. J. & de Teramond, G. F. The QCD Running Coupling. *Nucl. Phys.* **90,** 1. arXiv: 1604.08082 [hep-ph] (2016).

21. Cline, J. M. in *Advanced Concepts in Quantum Field Theory: With Exercises* 29–36 (Springer International Publishing, Cham, 2020). ISBN: 978-3-030-56168-0. https://doi.org/10.1007/978-3-030-56168-0_6.

22. Prosperi, G. M., Raciti, M. & Simolo, C. On the running coupling constant in QCD. *Prog. Part. Nucl. Phys.* **58,** 387–438. arXiv: hep-ph/0607209 (2007).

23. Buckley, A., White, C. & White, M. *Practical Collider Physics* ISBN: 978-0-7503-2444-1, 978-0-7503-2443-4, 978-0-7503-2442-7 (IOP, Dec. 2021).

24. Gross, D. J. & Wilczek, F. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys. Rev. Lett.* **30** (ed Taylor, J. C.) 1343–1346 (1973).

25. Politzer, H. D. Asymptotic Freedom: An Approach to Strong Interactions. *Phys. Rept.* **14,** 129–180 (1974).

26. Duke, D. W. & Roberts, R. G. Determinations of the QCD Strong Coupling alpha-s and the Scale Lambda (QCD). *Phys. Rept.* **120,** 275 (1985).

27. Kar, D. *Understanding the measurements and searches at the Large Hadron Collider* ISBN: 978-0-7503-2110-5 (IOP Publishing, Aug. 2019).

28. Taylor, R. E. Deep inelastic scattering: The Early years. *Rev. Mod. Phys.* **63,** 573–595 (1991).

29. Engelen, J. & Kooijman, P. Deep inelastic scattering at HERA: A review of experimental results in the light of quantum chromodynamics. *Prog. Part. Nucl. Phys.* **41,** 1–47 (1998).

30. Gribov, V. N. & Lipatov, L. N. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.* **15,** 438–450 (1972).

31. Dokshitzer, Y. L. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP* **46,** 641–653 (1977).

32. Altarelli, G. & Parisi, G. Asymptotic Freedom in Parton Language. *Nucl. Phys. B* **126,** 298–318 (1977).

33. Sjostrand, T. A Model for Initial State Parton Showers. *Phys. Lett. B* **157,** 321–325 (1985).

34. Marchesini, G. & Webber, B. R. Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation. *Nucl. Phys. B* **310,** 461–526 (1988).

35. Nagy, Z. & Soper, D. E. What is a parton shower? *Phys. Rev. D* **98,** 014034. arXiv: 1705.08093 [hep-ph] (2018).

36. Sudakov, V. V. Vertex parts at very high-energies in quantum electrodynamics. *Sov. Phys. JETP* **3,** 65–71 (1956).

37. Collins, J. C. Sudakov form-factors. *Adv. Ser. Direct. High Energy Phys.* **5,** 573–614. arXiv: hep-ph/0312336 (1989).

38. Nagy, Z. & Soper, D. E. Ordering variable for parton showers. *JHEP* **06,** 178. arXiv: 1401.6366 [hep-ph] (2014).

39. Gottschalk, T. D. BACKWARDS EVOLVED INITIAL STATE PARTON SHOWERS. *Nucl. Phys. B* **277,** 700–738 (1986).

40. Catani, S. *Soft gluon resummation: A Short review* in *32nd Rencontres de Moriond: QCD and High-Energy Hadronic Interactions* (1997), 331–336. arXiv: hep-ph/9709503.

41. Bellm, J. *et al.* Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C* **76,** 196. arXiv: 1512.01178 [hep-ph] (2016).

42. Sjöstrand, T., Mrenna, S. & Skands, P. Z. PYTHIA 6.4 physics and manual. *JHEP* **05,** 026. arXiv: hep-ph/0603175 (2006).

43. Bierlich, C. *et al.* A comprehensive guide to the physics and usage of PYTHIA 8.3. arXiv: 2203.11601 [hep-ph] (Mar. 2022).

44. Bothmann, E. *et al.* Event Generation with Sherpa 2.2. *SciPost Phys.* **7,** 034. arXiv: 1905.09127 [hep-ph] (2019).

45. Alioli, S., Nason, P., Oleari, C. & Re, E. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP* **06,** 043. arXiv: 1002.2581 [hep-ph] (2010).

46. Sjostrand, T. & van Zijl, M. A Multiple Interaction Model for the Event Structure in Hadron Collisions. *Phys. Rev. D* **36,** 2019 (1987).

47. Buckley, A. *et al.* General-purpose event generators for LHC physics. *Phys. Rept.* **504,** 145–233. arXiv: 1101.2599 [hep-ph] (2011).

48. Höche, S. *Introduction to parton-shower event generators* in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders* (2015), 235–295. arXiv: 1411.4085 [hep-ph].

49. Webber, B. R. A QCD Model for Jet Fragmentation Including Soft Gluon Interference. *Nucl. Phys. B* **238,** 492–528 (1984).

50. Artru, X. & Mennessier, G. String model and multiproduction. *Nucl. Phys. B* **70,** 93–115 (1974).

51. Field, R. D. & Feynman, R. P. A Parametrization of the Properties of Quark Jets. *Nucl. Phys. B* **136** (ed Brown, L. M.) 1 (1978).

52. Andersson, B., Gustafson, G., Ingelman, G. & Sjostrand, T. Parton Fragmentation and String Dynamics. *Phys. Rept.* **97,** 31–145 (1983).

53. Andersson, B. *The Lund model* ISBN: 978-0-521-01734-3, 978-0-521-42094-5, 978-0-511-88149-7 (Cambridge University Press, July 2005).

54. Marchesini, G. & Webber, B. R. Simulation of QCD Jets Including Soft Gluon Interference. *Nucl. Phys. B* **238,** 1–29 (1984).

55. Amati, D. & Veneziano, G. Preconfinement as a Property of Perturbative QCD. *Phys. Lett. B* **83,** 87–92 (1979).

56. Coleman, S. R. & Witten, E. Chiral Symmetry Breakdown in Large N Chromodynamics. *Phys. Rev. Lett.* **45,** 100 (1980).

57. Gustafson, G., Pettersson, U. & Zerwas, P. M. Jet Final States in W W Pair Production and Color Screening in the QCD Vacuum. *Phys. Lett. B* **209,** 90–94 (1988).

58. Sjostrand, T. & Khoze, V. A. On Color rearrangement in hadronic W+ W- events. *Z. Phys. C* **62,** 281–310. arXiv: hep-ph/9310242 (1994).

59. Mrenna, S. & Richardson, P. Matching matrix elements and parton showers with HERWIG and PYTHIA. *JHEP* **05,** 040. arXiv: hep-ph/0312274 (2004).

60. Alwall, J. *et al.* Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C* **53,** 473–500. arXiv: 0706.2569 [hep-ph] (2008).

61. Catani, S., Krauss, F., Kuhn, R. & Webber, B. R. QCD matrix elements + parton showers. *JHEP* **11,** 063. arXiv: hep-ph/0109231 (2001).

62. Krauss, F. Matrix elements and parton showers in hadronic interactions. *JHEP* **08,** 015. arXiv: hep-ph/0205283 (2002).

63. Mangano, M. L., Moretti, M. & Pittau, R. Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study. *Nucl. Phys. B* **632,** 343–362. arXiv: hep-ph/0108069 (2002).

64. Lonnblad, L. Correcting the color dipole cascade model with fixed order matrix elements. *JHEP* **05,** 046. arXiv: hep-ph/0112284 (2002).

65. Lavesson, N. & Lonnblad, L. Extending CKKW-merging to One-Loop Matrix Elements. *JHEP* **12,** 070. arXiv: 0811.2912 [hep-ph] (2008).

66. Frixione, S., Nason, P. & Oleari, C. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP* **11,** 070. arXiv: 0709.2092 [hep-ph] (2007).

67. Frixione, S. & Webber, B. R. Matching NLO QCD computations and parton shower simulations. *JHEP* **06,** 029. arXiv: hep-ph/0204244 (2002).

68. ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* **716,** 1–29. arXiv: 1207.7214 [hep-ex] (2012).

69. CMS Collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B* **716,** 30–61. arXiv: 1207.7235 [hep-ex] (2012).

70. Susskind, L. THE GAUGE HIERARCHY PROBLEM, TECHNICOLOR, SUPERSYMMETRY, AND ALL THAT. *Phys. Rept.* **104** (ed Mosher, A.) 181–193 (1984).

71. Fisher, P., Kayser, B. & McFarland, K. S. Neutrino mass and oscillation. *Ann. Rev. Nucl. Part. Sci.* **49,** 481–528. arXiv: hep-ph/9906244 (1999).

72. Cline, J. M. *Baryogenesis* in *Les Houches Summer School - Session 86: Particle Physics and Cosmology: The Fabric of Spacetime* (Sept. 2006). arXiv: hep-ph/0609145.

73. Peccei, R. D. The Strong CP problem and axions. *Lect. Notes Phys.* **741** (eds Kuster, M., Raffelt, G. & Beltran, B.) 3–17. arXiv: hep-ph/0607268 (2008).

74. Ade, P. A. R. *et al.* Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.* **571,** A16. arXiv: 1303.5076 [astro-ph.CO] (2014).

75. Zwicky, F. Die Rotverschiebung von extragalaktischen Nebeln. *Helv. Phys. Acta* **6,** 110–127 (1933).

76. Bertone, G. & Hooper, D. History of dark matter. *Rev. Mod. Phys.* **90,** 045002. arXiv: 1605.04909 [astro-ph.CO] (2018).

77. Baudis, L. The Search for Dark Matter. arXiv: 1801.08128 [astro-ph.CO] (Jan. 2018).

78. Rubin, V. C. & Ford Jr., W. K. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *Astrophys. J.* **159,** 379–403 (1970).

79. Bosma, A. *The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types* PhD thesis (University of Groningen, Netherlands, Mar. 1978).

80. Bosma, A. 21-cm line studies of spiral galaxies. 2. The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types. *Astron. J.* **86,** 1825 (1981).

81. Rubin, V. C., Burstein, D., Ford Jr., W. K. & Thonnard, N. Rotation velocities of 16 SA galaxies and a comparison of Sa, Sb, and SC rotation properties. *Astrophys. J.* **289,** 81 (1985).

82. Penzias, A. A. & Wilson, R. W. A Measurement of Excess Antenna Temperature at 4080 Mc/s. **142,** 419–421 (July 1965).

83. Dicke, R. H., Peebles, P. J. E., Roll, P. G. & Wilkinson, D. T. Cosmic Black-Body Radiation. **142,** 414–419 (July 1965).

84. Linde, A. D. *Particle physics and inflationary cosmology* arXiv: hep-th/0503203 (1990).

85. Ade, P. A. R. *et al.* Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.* **594,** A13. arXiv: 1502.01589 [astro-ph.CO] (2016).

86. Bertone, G., Hooper, D. & Silk, J. Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.* **405,** 279–390. arXiv: hep-ph/0404175 (2005).

87. Frenk, C. S. & White, S. D. M. Dark matter and cosmic structure. *Annalen Phys.* **524,** 507–534. arXiv: 1210.0544 [astro-ph.CO] (2012).

88. Steigman, G. & Turner, M. S. Cosmological Constraints on the Properties of Weakly Interacting Massive Particles. *Nucl. Phys. B* **253,** 375–386 (1985).

89. Wess, J. & Zumino, B. Supergauge Transformations in Four-Dimensions. *Nucl. Phys. B* **70** (eds Salam, A. & Sezgin, E.) 39–50 (1974).

90. Martin, S. P. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.* **18** (ed Kane, G. L.) 1–98. arXiv: hep-ph/9709356 (1998).

91. Peccei, R. D. & Quinn, H. R. CP Conservation in the Presence of Instantons. *Phys. Rev. Lett.* **38,** 1440–1443 (1977).

92. Kim, J. E. & Carosi, G. Axions and the Strong CP Problem. *Rev. Mod. Phys.* **82.** [Erratum: Rev.Mod.Phys. 91, 049902 (2019)], 557–602. arXiv: 0807.3125 [hep-ph] (2010).

93. Khan Cantlay, B. & Wechakama, M. Constraints on dark matter annihilation with electron spectrum from VERITAS. *J. Phys. Conf. Ser.* **1380** (eds Laosiritaworn, Y., Tepnual, T., Kessaratikoon, P. & Thountom, S.) 012071 (2019).

94. Rinchiuso, L., Moulin, E., Armand, C. & Poireau, V. Dark Matter search with H.E.S.S. towards ultra-faint dwarf nearby DES satellites of the Milky Way. *PoS* **ICRC2019,** 542. arXiv: 1908.04311 [astro-ph.HE] (2020).

95. Acciari, V. A. *et al.* A search for dark matter in Triangulum II with the MAGIC telescopes. *Phys. Dark Univ.* **28,** 100529. arXiv: 2003.05260 [astro-ph.HE] (2020).

96. Abazajian, K. N., Horiuchi, S., Kaplinghat, M., Keeley, R. E. & Macias, O. Strong constraints on thermal relic dark matter from Fermi-LAT observations of the Galactic Center. *Phys. Rev. D* **102,** 043012. arXiv: 2003.10416 [hep-ph] (2020).

97. Ruchayskiy, O. *et al.* Searching for decaying dark matter in deep XMM–Newton observation of the Draco dwarf spheroidal. *Mon. Not. Roy. Astron. Soc.* **460,** 1390–1398. arXiv: 1512.07217 [astro-ph.HE] (2016).

98. Neronov, A., Malyshev, D. & Eckert, D. Decaying dark matter search with NuSTAR deep sky observations. *Phys. Rev. D* **94,** 123504. arXiv: 1607.07328 [astro-ph.HE] (2016).

99. Sekiya, N., Yamasaki, N. Y. & Mitsuda, K. A Search for a keV Signature of Radiatively Decaying Dark Matter with Suzaku XIS Observations of the X-ray Diffuse Background. arXiv: 1504.02826 [astro-ph.HE] (Apr. 2015).

100. Abbasi, R. *et al.* A Search for Neutrinos from Decaying Dark Matter in Galaxy Clusters and Galaxies with IceCube. *PoS* **ICRC2021,** 506. arXiv: 2107.11527 [astro-ph.HE] (2021).

101. Albert, A. *et al.* Search for dark matter towards the Galactic Centre with 11 years of ANTARES data. *Phys. Lett. B* **805,** 135439. arXiv: 1912.05296 [astro-ph.HE] (2020).

102. Avrorin, A. D. *et al.* Dark matter constraints from an observation of dSphs and the LMC with the Baikal NT200. *J. Exp. Theor. Phys.* **125,** 80–90. arXiv: 1612.03836 [astro-ph.HE] (2017).

103. Barinov, V. V. *et al.* Results from the Baksan Experiment on Sterile Transitions (BEST). *Phys. Rev. Lett.* **128,** 232501. arXiv: 2109.11482 [nucl-ex] (2022).

104. Takeuchi, Y. Recent results and future prospects of Super-Kamiokande. *Nucl. Instrum. Meth. A* **952,** 161634 (2020).

105. Abeysekara, A. U. *et al.* A Search for Dark Matter in the Galactic Halo with HAWC. *JCAP* **02,** 049. arXiv: 1710.10288 [astro-ph.HE] (2018).

106. Wechakama, M. & Khan Cantlay, B. Upper limits on dark matter annihilation with the teraelectronvolt cosmic ray spectrum of electrons and positrons from DAMPE. *J. Phys. Conf. Ser.* **1380** (eds Laosiritaworn, Y., Tepnual, T., Kessaratikoon, P. & Thountom, S.) 012144 (2019).

107. Adriani, O. *et al.* Extended Measurement of the Cosmic-Ray Electron and Positron Spectrum from 11 GeV to 4.8 TeV with the Calorimetric Electron Telescope on the International Space Station. *Phys. Rev. Lett.* **120,** 261102. arXiv: 1806.09728 [astro-ph.HE] (2018).

108. Aguilar, M. *et al.* The Alpha Magnetic Spectrometer (AMS) on the international space station: Part II – Results from the first seven years. *Phys. Rept.* **894,** 1–116 (2021).

109. Aartsen, M. G. *et al.* IceCube-Gen2: the window to the extreme Universe. *J. Phys. G* **48,** 060501. arXiv: 2008.04323 [astro-ph.HE] (2021).

110. Adrian-Martinez, S. *et al.* Letter of intent for KM3NeT 2.0. *J. Phys. G* **43,** 084001. arXiv: 1601.07459 [astro-ph.IM] (2016).

111. Actis, M. *et al.* Design concepts for the Cherenkov Telescope Array CTA: An advanced facility for ground-based high-energy gamma-ray astronomy. *Exper. Astron.* **32,** 193–316. arXiv: 1008.3703 [astro-ph.IM] (2011).

112. Shoibonov, B. Baikal-GVD - the Next Generation Neutrino Telescope in Lake Baikal. *J. Phys. Conf. Ser.* **1263** (eds Ivanov, I. & Kazarina, Y.) 012005 (2019).

113. Abe, K. *et al.* Letter of Intent: The Hyper-Kamiokande Experiment — Detector Design and Physics Potential —. arXiv: 1109.3262 [hep-ex] (Sept. 2011).

114. Pérez de los Heros, C. Status, Challenges and Directions in Indirect Dark Matter Searches. *Symmetry* **12,** 1648. arXiv: 2008.11561 [astro-ph.HE] (2020).

115. Goodman, M. W. & Witten, E. Detectability of Certain Dark Matter Candidates. *Phys. Rev. D* **31** (ed Srednicki, M. A.) 3059 (1985).

116. Cox, P., Melia, T. & Rajendran, S. Dark matter phonon coupling. *Phys. Rev. D* **100,** 055011. arXiv: 1905.05575 [hep-ph] (2019).

117. Abdelhameed, A. H. *et al.* First results from the CRESST-III low-mass dark matter program. *Phys. Rev. D* **100,** 102002. arXiv: 1904.00498 [astro-ph.CO] (2019).

118. Aprile, E. *et al.* XENON100 Dark Matter Results from a Combination of 477 Live Days. *Phys. Rev. D* **94,** 122001. arXiv: 1609.06154 [astro-ph.CO] (2016).

119. Bernabei, R. *et al.* First model independent results from DAMA/LIBRA-phase2. *Nucl. Phys. Atom. Energy* **19,** 307–325. arXiv: 1805.10486 [hep-ex] (2018).

120. Lin, T. Dark matter models and direct detection. *PoS* **333,** 009. arXiv: 1904.07915 [hep-ph] (2019).

121. Billard, J. *et al.* Direct detection of dark matter–APPEC committee report*. *Rept. Prog. Phys.* **85,** 056201. arXiv: 2104.07634 [hep-ex] (2022).

122. Goldberg, H. & Hall, L. J. A New Candidate for Dark Matter. *Phys. Lett. B* **174** (ed Tran Thanh Van, J.) 151 (1986).

123. Strassler, M. J. & Zurek, K. M. Echoes of a hidden valley at hadron colliders. *Phys. Lett. B* **651,** 374–379. arXiv: hep-ph/0604261 (2007).

124. Pospelov, M., Ritz, A. & Voloshin, M. B. Secluded WIMP Dark Matter. *Phys. Lett. B* **662,** 53–61. arXiv: 0711.4866 [hep-ph] (2008).

125. Feng, J. L. & Kumar, J. The WIMPless Miracle: Dark-Matter Particles without Weak-Scale Masses or Weak Interactions. *Phys. Rev. Lett.* **101,** 231301. arXiv: 0803.4196 [hep-ph] (2008).

126. Arkani-Hamed, N., Finkbeiner, D. P., Slatyer, T. R. & Weiner, N. A Theory of Dark Matter. *Phys. Rev. D* **79,** 015014. arXiv: 0810.0713 [hep-ph] (2009).

127. Patt, B. & Wilczek, F. Higgs-field portal into hidden sectors. arXiv: hep-ph/0605188 (May 2006).

128. De Simone, A. & Jacques, T. Simplified models vs. effective field theory approaches in dark matter searches. *Eur. Phys. J. C* **76,** 367. arXiv: 1603.08002 [hep-ph] (2016).

129. Chetyrkin, K. G., Kniehl, B. A. & Steinhauser, M. Decoupling relations to O (alpha-s^3) and their connection to low-energy theorems. *Nucl. Phys. B* **510,** 61–87. arXiv: hep-ph/9708255 (1998).

130. Arina, C., Fuks, B. & Mantani, L. A universal framework for t-channel dark matter models. *Eur. Phys. J. C* **80,** 409. arXiv: 2001.05024 [hep-ph] (2020).

131. Bai, Y. & Berger, J. Fermion Portal Dark Matter. *JHEP* **11,** 171. arXiv: 1308.0612 [hep-ph] (2013).

132. Bai, Y. & Schwaller, P. Scale of dark QCD. *Phys. Rev. D* **89,** 063522. arXiv: 1306.4676 [hep-ph] (2014).

133. Buchmueller, O., Dolan, M. J. & McCabe, C. Beyond Effective Field Theory for Dark Matter Searches at the LHC. *JHEP* **01,** 025. arXiv: 1308.6799 [hep-ph] (2014).

134. Dreiner, H., Schmeier, D. & Tattersall, J. Contact Interactions Probe Effective Dark Matter Models at the LHC. *EPL* **102,** 51001. arXiv: 1303.3348 [hep-ph] (2013).

135. Harris, P., Khoze, V. V., Spannowsky, M. & Williams, C. Constraining Dark Sectors at Colliders: Beyond the Effective Theory Approach. *Phys. Rev. D* **91,** 055009. arXiv: 1411.0535 [hep-ph] (2015).

136. Cohen, T., Lisanti, M. & Lou, H. K. Semivisible Jets: Dark Matter Undercover at the LHC. *Phys. Rev. Lett.* **115,** 171804. arXiv: 1503.00009 [hep-ph] (2015).

137. Liew, S. P., Papucci, M., Vichi, A. & Zurek, K. M. Mono-X Versus Direct Searches: Simplified Models for Dark Matter at the LHC. *JHEP* **06,** 082. arXiv: 1612.00219 [hep-ph] (2017).

138. Englert, C., McCullough, M. & Spannowsky, M. S-Channel Dark Matter Simplified Models and Unitarity. *Phys. Dark Univ.* **14,** 48–56. arXiv: 1604.07975 [hep-ph] (2016).

139. Bernreuther, E., Kahlhoefer, F., Krämer, M. & Tunney, P. Strongly interacting dark sectors in the early Universe and at the LHC through a simplified portal. *JHEP* **01,** 162. arXiv: 1907.04346 [hep-ph] (2020).

140. An, H., Wang, L.-T. & Zhang, H. Dark matter with t-channel mediator: a simple step beyond contact interaction. *Phys. Rev. D* **89,** 115014. arXiv: 1308.0592 [hep-ph] (2014).

141. Chang, S., Edezhath, R., Hutchinson, J. & Luty, M. Effective WIMPs. *Phys. Rev. D* **89,** 015011. arXiv: 1307.8120 [hep-ph] (2014).

142. Papucci, M., Vichi, A. & Zurek, K. M. Monojet versus the rest of the world I: t-channel models. *JHEP* **11,** 024. arXiv: 1402.2285 [hep-ph] (2014).

143. Cohen, T., Lisanti, M., Lou, H. K. & Mishra-Sharma, S. LHC Searches for Dark Sector Showers. *JHEP* **11,** 196. arXiv: 1707.05326 [hep-ph] (2017).

144. Han, T., Si, Z., Zurek, K. M. & Strassler, M. J. Phenomenology of hidden valleys at hadron colliders. *JHEP* **07,** 008. arXiv: 0712.2041 [hep-ph] (2008).

145. Carloni, L. & Sjostrand, T. Visible Effects of Invisible Hidden Valley Radiation. *JHEP* **09,** 105. arXiv: 1006.2911 [hep-ph] (2010).

146. Seth, M. S. *A first study of Hidden Valley models at the LHC* Bachelor thesis (Lund Observ., 2011). arXiv: 1106.2064 [hep-ph].

147. Schwaller, P., Stolarski, D. & Weiler, A. Emerging Jets. *JHEP* **05,** 059. arXiv: 1502.05409 [hep-ph] (2015).

148. Park, M. & Zhang, M. Tagging a jet from a dark sector with Jet-substructures at colliders. *Phys. Rev. D* **100,** 115009. arXiv: 1712.09279 [hep-ph] (2019).

149. Beauchesne, H., Bertuzzo, E. & Grilli Di Cortona, G. Dark matter in Hidden Valley models with stable and unstable light dark mesons. *JHEP* **04,** 118. arXiv: 1809.10152 [hep-ph] (2019).

150. Mies, H., Scherb, C. & Schwaller, P. Collider constraints on dark mediators. *JHEP* **04,** 049. arXiv: 2011.13990 [hep-ph] (2021).

151. De Melis, C. *The CERN accelerator complex.* https://cds.cern.ch/record/2119882. General Photo. 2016.

152. LHC Machine. *JINST* **3** (eds Evans, L. & Bryant, P.) S08001 (2008).

153. ATLAS Collaboration. *Public ATLAS Luminosity Results for Run-2 of the LHC* https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. General Photo. 2015.

154. Levin, E. *The Pomeron: Yesterday, today and tomorrow* in *3rd Gleb Wataghin School on High-energy Phemomenology* (Sept. 1994), 0158–261. arXiv: hep-ph/9503399.

155. ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST* **3,** S08003 (2008).

156. Pequenao, J. *Computer generated image of the whole ATLAS detector* 2008. https://cds.cern.ch/record/1095924.

157. Abbott, B. *et al.* Production and integration of the ATLAS Insertable B-Layer. *JINST* **13,** T05008. arXiv: 1803.00844 [physics.ins-det] (2018).

158. ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report* ATLAS-TDR-19; CERN-LHCC-2010-013 (2010). https://cds.cern.ch/record/1291633. Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009 (2012). https://cds.cern.ch/record/1451888.

159. Aad, G. *et al.* Operation and performance of the ATLAS semiconductor tracker. *JINST* **9,** P08009. arXiv: 1404.7473 [hep-ex] (2014).

160. ATLAS Collaboration. ATLAS liquid argon calorimeter: Technical design report (Dec. 1996).

161. Pequenao, J. *Computer generated image of the ATLAS inner detector* 2008. https://cds.cern.ch/record/1095926.

162. *ATLAS liquid-argon calorimeter: Technical Design Report* https://cds.cern.ch/record/331061 (CERN, Geneva, 1996).

163. Pequenao, J. *Computer Generated image of the ATLAS calorimeter* 2008. https://cds.cern.ch/record/1095927.

164. Airapetian, A. *et al.* ATLAS calorimeter performance Technical Design Report (Dec. 1996).

165. Pequenao, J. *Computer generated image of the ATLAS Muons subsystem* 2008. https://cds.cern.ch/record/1095929.

166. *Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report* (Mar. 2018). arXiv: 1803.07977 [hep-ph].

167. ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C* **77,** 466. arXiv: 1703.10485 [hep-ex] (2017).

168. Huth, J. E. *et al. Toward a standardization of jet definitions* in *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90)* (Dec. 1990), 0134–136.

169. Sterman, G. F. & Weinberg, S. Jets from Quantum Chromodynamics. *Phys. Rev. Lett.* **39,** 1436 (1977).

170. Salam, G. P. & Soyez, G. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP* **05,** 086. arXiv: 0704.0292 [hep-ph] (2007).

171. Cacciari, M., Salam, G. P. & Soyez, G. The anti-k_t jet clustering algorithm. *JHEP* **04,** 063. arXiv: 0802.1189 [hep-ph] (2008).

172. Dokshitzer, Y. L., Leder, G. D., Moretti, S. & Webber, B. R. Better jet clustering algorithms. *JHEP* **08,** 001. arXiv: hep-ph/9707323 (1997).

173. Wobisch, M. & Wengler, T. *Hadronization corrections to jet cross-sections in deep inelastic scattering* in *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)* (Apr. 1998), 270–279. arXiv: hep-ph/9907280.

174. Catani, S., Dokshitzer, Y. L., Olsson, M., Turnock, G. & Webber, B. R. New clustering algorithm for multi - jet cross-sections in e+ e- annihilation. *Phys. Lett. B* **269,** 432–438 (1991).

175. Ellis, S. D. & Soper, D. E. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D* **48,** 3160–3166. arXiv: hep-ph/9305266 (1993).

176. Cacciari, M., Salam, G. P. & Soyez, G. FastJet user manual. *Eur. Phys. J. C* **72,** 1896. arXiv: 1111.6097 [hep-ph] (2012).

177. ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. *Eur. Phys. J. C* **74,** 3071. arXiv: 1407.5063 [hep-ex] (2014).

178. ATLAS Collaboration. Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in *pp* collisions at $\sqrt{s}$=8 TeV. *Eur. Phys. J. C* **75,** 303. arXiv: 1412.7086 [hep-ex] (2015).

179. ATLAS Collaboration. Performance of algorithms that reconstruct missing transverse momentum in $\sqrt{s} = 8$ TeV proton-proton collisions in the ATLAS detector. *Eur. Phys. J. C* **77,** 241. arXiv: 1609.09324 [hep-ex] (2017).

180. ATLAS Collaboration. Performance of the ATLAS Trigger System in 2015. *Eur. Phys. J. C* **77,** 317. arXiv: 1611.09661 [hep-ex] (2017).

181. ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C* **70,** 823–874. arXiv: 1005.4568 [physics.ins-det] (2010).

182. Agostinelli, S. *et al.* GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A* **506,** 250–303 (2003).

183. Boveia, A. & Doglioni, C. Dark Matter Searches at Colliders. *Ann. Rev. Nucl. Part. Sci.* **68,** 429–459. arXiv: 1810.12238 [hep-ex] (2018).

184. Ward, E. *LHC detector transverse cross-section* General Photo. 2019. https://cds.cern.ch/record/2665178.

185. ATLAS Collaboration. Search for new phenomena in events with an energetic jet and missing transverse momentum in *pp* collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector. *Phys. Rev. D* **103,** 112006. arXiv: 2102.10874 [hep-ex] (2021).

186. CMS Collaboration. Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP* **11,** 153. arXiv: 2107.13021 [hep-ex] (2021).

187. ATLAS Collaboration. Search for dark matter in association with an energetic photon in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP* **02,** 226. arXiv: 2011.05259 [hep-ex] (2021).

188. CMS Collaboration. Search for new physics in final states with a single photon and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP* **02,** 074. arXiv: 1810.00196 [hep-ex] (2019).

189. CMS Collaboration. Search for dark matter produced in association with a leptonically decaying Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C* **81.** [Erratum: Eur.Phys.J.C 81, 333 (2021)], 13. arXiv: 2008.04735 [hep-ex] (2021).

190. ATLAS Collaboration. Search for associated production of a Z boson with an invisibly decaying Higgs boson or dark matter candidates at $\sqrt{s}$ =13 TeV with the ATLAS detector. *Phys. Lett. B* **829,** 137066. arXiv: 2111.08372 [hep-ex] (2022).

191. ATLAS Collaboration. Search for invisible Higgs-boson decays in events with vector-boson fusion signatures using 139 fb^{-1} of proton-proton data recorded by the ATLAS experiment. *JHEP* **08,** 104. arXiv: 2202.07953 [hep-ex] (2022).

192. CMS Collaboration. Search for invisible decays of the Higgs boson produced via vector boson fusion in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. D* **105,** 092007. arXiv: 2201.11585 [hep-ex] (2022).

193. ATLAS Collaboration. Search for dark matter in events with missing transverse momentum and a Higgs boson decaying into two photons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP* **10,** 013. arXiv: 2104.13240 [hep-ex] (2021).

194. Search for dark matter in final states with a Higgs boson decaying to a pair of b-jets and missing transverse momentum at the HL-LHC (2022).

195. ATLAS Collaboration. *Search for non-resonant production of semi-visible jets using Run 2 data in ATLAS* tech. rep. (CERN, Geneva, 2022). https://cds.cern.ch/record/2815284.

196. ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C* **77,** 466. arXiv: 1703.10485 [hep-ex] (2017).

197. ATLAS Collaboration. Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D* **96,** 072002. arXiv: 1703.09665 [hep-ex] (2017).

198. ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *Eur. Phys. J. C* **76,** 581. arXiv: 1510.03823 [hep-ex] (2016).

199. ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *Eur. Phys. J. C* **76,** 581. arXiv: 1510.03823 [hep-ex] (2016).

200. Nachman, B., Nef, P., Schwartzman, A., Swiatlowski, M. & Wanotayaroj, C. Jets from Jets: Re-clustering as a tool for large radius jet reconstruction and grooming at the LHC. *JHEP* **02,** 075. arXiv: 1407.2922 [hep-ph] (2015).

201. ATLAS Collaboration. *Jet reclustering and close-by effects in ATLAS Run 2* ATLAS-CONF-2017-062. 2017. https://cds.cern.ch/record/2275649.

202. ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run* ATL-PHYS-PUB-2017-013. 2017. https://cds.cern.ch/record/2273281.

203. ATLAS Collaboration. Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP* **03,** 145. arXiv: 1910.08447 [hep-ex] (2020).

204. ATLAS Collaboration. Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13\,\text{TeV}$. *JHEP* **08,** 089. arXiv: 1805.01845 [hep-ex] (2018).

205. ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C* **78,** 903. arXiv: 1802.08168 [hep-ex] (2018).

206. ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data. *JINST* **14,** P12006. arXiv: 1908.00005 [hep-ex] (2019).

207. ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C* **76,** 292. arXiv: 1603.05598 [hep-ex] (2016).

208. ATLAS Collaboration. *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment* ATL-PHYS-PUB-2019-033. 2019. https://cds.cern.ch/record/2688062.

209. Alwall, J. *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP* **07,** 079. arXiv: 1405.0301 [hep-ph] (2014).

210. *DMSimp_tchannel* https://github.com/smsharma/SemivisibleJets/tree/master/DMsimp_tchannel.

211. Sjöstrand, T., Mrenna, S. & Skands, P. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.* **178,** 852–867. arXiv: 0710.3820 [hep-ph] (2008).

212. Sjöstrand, T. *et al.* An introduction to PYTHIA 8.2. *Comput. Phys. Commun.* **191,** 159. arXiv: 1410.3012 [hep-ph] (2015).

213. Ball, R. D. *et al.* Parton distributions for the LHC run II. *JHEP* **04,** 040. arXiv: 1410.8849 [hep-ph] (2015).

214. ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data* ATL-PHYS-PUB-2014-021. 2014. https://cds.cern.ch/record/1966419.

215. Carloni, L., Rathsman, J. & Sjostrand, T. Discerning Secluded Sector gauge structures. *JHEP* **04,** 091. arXiv: 1102.3795 [hep-ph] (2011).

216. Bothmann, E. *et al.* Event generation with Sherpa 2.2. *SciPost Phys.* **7,** 034. arXiv: 1905.09127 [hep-ph] (2019).

217. Schumann, S. & Krauss, F. A parton shower algorithm based on Catani–Seymour dipole factorisation. *JHEP* **03,** 038. arXiv: 0709.1027 [hep-ph] (2008).

218. Nason, P. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP* **11,** 040. arXiv: hep-ph/0409146 (2004).

219. ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C* **70,** 823. arXiv: 1005.4568 [physics.ins-det] (2010).

220. GEANT4 Collaboration, Agostinelli, S., *et al.* GEANT4 – a simulation toolkit. *Nucl. Instrum. Meth. A* **506,** 250 (2003).

221. ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model* ATL-PHYS-PUB-2016-017. 2016. https://cds.cern.ch/record/2206965.

222. ATLAS Collaboration. Performance of the missing transverse momentum triggers for the ATLAS detector during Run-2 data taking. *JHEP* **08,** 080. arXiv: 2005.09554 [hep-ex] (2020).

223. ATLAS Collaboration. Search for squarks and gluinos in final states with jets and missing transverse momentum using 139 fb^{-1} of $\sqrt{s}$ =13 TeV pp collision data with the ATLAS detector. *JHEP* **02,** 143. arXiv: 2010.14293 [hep-ex] (2021).

224. ATLAS Collaboration. Characterisation and mitigation of beam-induced backgrounds observed in the ATLAS detector during the 2011 proton–proton run. *JINST* **8,** P07004. arXiv: 1303.0223 [hep-ex] (2013).

225. ATLAS Collaboration. *Selection of jets produced in* 13 *TeV proton–proton collisions with the ATLAS detector* ATLAS-CONF-2015-029. 2015. https://cds.cern.ch/record/2037702.

226. ATLAS Collaboration. Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C* **81,** 689. arXiv: 2007.02645 [hep-ex] (2021).

227. ATLAS Collaboration. Measurement of hard double-parton interactions in $W(\rightarrow l\nu)$+ 2 jet events at $\sqrt{s}$=7 TeV with the ATLAS detector. *New J. Phys.* **15,** 033038. arXiv: 1301.6872 [hep-ex] (2013).

228. ATLAS Collaboration. Measurement of the $t\bar{t}$ production cross-section and lepton differential distributions in $e\mu$ dilepton events from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C* **80,** 528. arXiv: 1910.08819 [hep-ex] (2020).

229. ATLAS Collaboration. Measurements of top-quark pair differential and double-differential cross-sections in the ℓ+jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *Eur. Phys. J. C* **79.** [Erratum: Eur.Phys.J.C 80, 1092 (2020)], 1028. arXiv: 1908.07305 [hep-ex] (2019).

230. Frixione, S., Laenen, E., Motylinski, P., White, C. & Webber, B. R. Single-top hadroproduction in association with a W boson. *JHEP* **07,** 029. arXiv: 0805.3067 [hep-ph] (2008).

231. ATLAS Collaboration. Simulation of top quark production for the ATLAS experiment at sqrt(s) = 13 TeV (Jan. 2016).

232. ATLAS Collaboration. Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector. *JHEP* **12,** 024. arXiv: 1708.03299 [hep-ex] (2017).

233. ATLAS Collaboration. Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C* **81,** 689. arXiv: 2007.02645 [hep-ex] (2020).

234. ATLAS Collaboration. Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data. *Eur. Phys. J. C* **74,** 3130. arXiv: 1407.3935 [hep-ex] (2014).

235. ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C* **81,** 578. arXiv: 2012.00578 [hep-ex] (2021).

236. Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC (June 2019).

237. Cowan, G., Cranmer, K., Gross, E. & Vitells, O. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C* **71,** 1554. arXiv: 1007.1727 (2011). Erratum: *Eur. Phys. J. C* **73,** 2501 (2013).

238. CMS Collaboration. Search for resonant production of strongly coupled dark matter in proton–proton collisions at 13 TeV. arXiv: 2112.11125 [hep-ex] (2021).

239. Bierlich, C. *et al.* Robust Independent Validation of Experiment and Theory: Rivet version 3. *SciPost Phys.* **8,** 026. arXiv: 1912.05451 [hep-ph] (2020).

240. Cacciari, M., Salam, G. P. & Soyez, G. FastJet User Manual. *Eur. Phys. J.* **C72,** 1896. arXiv: 1111.6097 [hep-ph] (2012).

241. Skands, P., Carrazza, S. & Rojo, J. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C* **74,** 3024. arXiv: 1404.5630 [hep-ph] (2014).

242. Kar, D. & Sinha, S. Exploring jet substructure in semi-visible jets. *SciPost Phys.* **10,** 084. arXiv: 2007.11597 [hep-ph] (2021).

243. Krohn, D., Thaler, J. & Wang, L.-T. Jet Trimming. *JHEP* **02,** 084. arXiv: 0912.1342 [hep-ph] (2010).

244. Kogler, R. *et al.* Jet Substructure at the Large Hadron Collider: Experimental Review. *Rev. Mod. Phys.* **91,** 045003. arXiv: 1803.06991 [hep-ex] (2019).

245. Marzani, S., Soyez, G. & Spannowsky, M. *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology* arXiv: 1901.10342 [hep-ph] (Springer, 2019).

246. Cohen, T., Doss, J. & Freytsis, M. Jet Substructure from Dark Sector Showers. *arXiv:2004.00631.* arXiv: 2004.00631 [hep-ph] (Apr. 2020).

247. Gras, P. *et al.* Systematics of quark/gluon tagging. *JHEP* **07,** 091. arXiv: 1704.03878 [hep-ph] (2017).

248. Larkoski, A. J., Marzani, S., Soyez, G. & Thaler, J. Soft Drop. *JHEP* **05,** 146. arXiv: 1402.2657 [hep-ph] (2014).

249. ATLAS Collaboration. Measurement of soft-drop jet observables in pp collisions with the ATLAS detector at $\sqrt{s}$ =13 TeV. *Phys. Rev. D* **101,** 052007. arXiv: 1912.09837 [hep-ex] (2020).

250. Thaler, J. & Van Tilburg, K. Identifying Boosted Objects with N-subjettiness. *JHEP* **03,** 015. arXiv: 1011.2268 [hep-ph] (2011).

251. Larkoski, A. J., Salam, G. P. & Thaler, J. Energy Correlation Functions for Jet Substructure. *JHEP* **06,** 108. arXiv: 1305.0007 [hep-ph] (2013).

252. Larkoski, A. J., Moult, I. & Neill, D. Analytic boosted boson discrimination. *JHEP* **05,** 117. arXiv: 1507.03018 [hep-ph] (2016).

253. Larkoski, A. J., Thaler, J. & Waalewijn, W. J. Gaining (Mutual) Information about Quark/Gluon Discrimination. *JHEP* **11,** 129. arXiv: 1408.3122 [hep-ph] (2014).

254. Larkoski, A. J., Moult, I. & Neill, D. Power counting to better jet observables. *JHEP* **12,** 009. arXiv: 1409.6298 [hep-ph] (2014).

255. Moult, I., Necib, L. & Thaler, J. New Angles on Energy Correlation Functions. *JHEP* **12,** 153. arXiv: 1609.07483 [hep-ph] (2016).

256. Buckley, A., Kar, D. & NordstrÃ¶m, K. Fast simulation of detector effects in Rivet. *SciPost Phys.* **8,** 025. arXiv: 1910.01637 [hep-ph] (2020).

257. Buckley, A., Kar, D. & Sinha, S. *Towards better discrimination and improved modelling of dark-sector showers* in *51st International Symposium on Multiparticle Dynamics* (Sept. 2022). arXiv: 2209.14964 [hep-ph].

258. Komiske, P. T., Metodiev, E. M. & Thaler, J. Energy flow polynomials: A complete linear basis for jet substructure. *JHEP* **04,** 013. arXiv: 1712.07124 [hep-ph] (2018).

259. Faucett, T., Hsu, S.-C. & Whiteson, D. Learning to Identify Semi-Visible Jets. arXiv: 2208.10062 [hep-ph] (Aug. 2022).

260. Kar, D. & Sinha, S. 2B or not 2B, a study of bottom-quark-philic semi-visible jets. arXiv: 2207.01885 [hep-ph] (July 2022).

261. Strassler, M. J. On the Phenomenology of Hidden Valleys with Heavy Flavor. arXiv: 0806.2385 [hep-ph] (June 2008).

262. Renner, S. & Schwaller, P. A flavoured dark sector. *JHEP* **08,** 052. arXiv: 1803.08080 [hep-ph] (2018).

263. ATLAS Collaboration. Performance of the ATLAS trigger system in 2015. *Eur. Phys. J. C* **77,** 317. arXiv: 1611.09661 [hep-ex] (2017).

264. ATLAS Collaboration. Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions. *Eur. Phys. J. C* **81,** 334. arXiv: 2009.04986 [hep-ex] (2020).

265. ATLAS Collaboration. Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV. *Eur. Phys. J. C* **79,** 836. arXiv: 1906.11005 [hep-ex] (2019).

266. Krohn, D., Thaler, J. & Wang, L.-T. Jets with Variable R. *JHEP* **06,** 059. arXiv: 0903.0392 [hep-ph] (2009).

267. ATLAS Collaboration. Search for dark matter produced in association with a Standard Model Higgs boson decaying into b-quarks using the full Run 2 dataset from the ATLAS detector. *JHEP* **11,** 209. arXiv: 2108.13391 [hep-ex] (2021).

268. Bernreuther, E., Finke, T., Kahlhoefer, F., KrÃ¤mer, M. & MÃ¼ck, A. Casting a graph net to catch dark showers. *arXiv:2006.08639*. arXiv: 2006.08639 [hep-ph] (June 2020).

269. Butter, A., Kasieczka, G., Plehn, T. & Russell, M. Deep-learned Top Tagging with a Lorentz Layer. *SciPost Phys.* **5,** 028. arXiv: 1707.08966 [hep-ph] (2018).

270. Andreassen, A., Feige, I., Frye, C. & Schwartz, M. D. Binary JUNIPR: an interpretable probabilistic model for discrimination. *Phys. Rev. Lett.* **123,** 182001. arXiv: 1906.10137 [hep-ph] (2019).

271. Chen, Y.-C. J., Chiang, C.-W., Cottin, G. & Shih, D. Boosted W and Z tagging with jet charge and deep learning. *Phys. Rev. D* **101,** 053001. arXiv: 1908.08256 [hep-ph] (2020).

272. Moore, L., Nordström, K., Varma, S. & Fairbairn, M. Reports of My Demise Are Greatly Exaggerated: N-subjettiness Taggers Take On Jet Images. *SciPost Phys.* **7,** 036. arXiv: 1807.04769 [hep-ph] (2019).

273. Komiske, P. T., Metodiev, E. M. & Thaler, J. Energy Flow Networks: Deep Sets for Particle Jets. *JHEP* **01,** 121. arXiv: 1810.05165 [hep-ph] (2019).

274. Chakraborty, A., Lim, S. H., Nojiri, M. M. & Takeuchi, M. Neural Network-based Top Tagger with Two-Point Energy Correlations and Geometry of Soft Emissions. *JHEP* **20,** 111. arXiv: 2003.11787 [hep-ph] (2020).

275. ATLAS Collaboration. *Proposal for particle-level object and observable definitions for use in physics measurements at the LHC* ATL-PHYS-PUB-2015-013. 2015. https://cds.cern.ch/record/2022743.

276. Buckley, A. *et al.* The HepMC3 event record library for Monte Carlo event generators. *Comput. Phys. Commun.* **260,** 107310. arXiv: 1912.08005 [hep-ph] (2021).

277. *Monte Carlo Particle Numbering Scheme* https://pdg.lbl.gov/2019/reviews/rpp2019-rev-monte-carlo-numbering.pdf. Accessed: 2020-11-10.

Appendix A

Appendices

A.1 Use of reclustered jets in ATLAS SVJ analysis

This appendix section highlights the several studies conducted to check whether reclustering can be a viable option when reconstructing the jets, and the reason why it was not ultimately selected.

A.1.1 Particle level studies

In order to understand the final state topology better, the relevant objects are shown in the $\eta - \phi$ plane for six representative events, as shown in Fig A.2. The large hollow blue circles represent the trimmed large-radius jets, the filled cyan circles represent anti-k_t jets with R = 0.4, the black points represent dark hadrons, and the red line represents the direction of missing transverse momentum. These particle level plots show that large radius jets better contain the semi-visible jets. However, as can be seen from the last three diagrams, jets with radius larger than $R = 1.0$ is optimal for enclosing the clusters of energy coming from a SVJ. Therefore it was decided to move to reclustered anti-k_t jets with R=1.5, a concept first proposed in [200]. Here, particle level studies are shown, and in the next subsection, detector level considerations will be discussed.

In Fig A.1, the area normalised H_{T} distributions are shown to aid shape comparisons. In Fig. A.3, kinematic distributions of reclustered anti-k_t $R = 1.5$ jet are shown. Finally, in Fig. A.4, the average overlap between all reclustered jets is shown, for three different radii, $R = 1.0$, $R = 1.2$ and $R = 1.5$. The average overlap is calculated as the sum of ΔR between all pairs of overlapping jets, divided by number of overlaps.

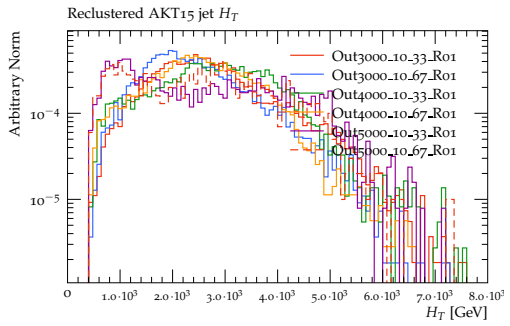


Figure A.1: Particle level distribution of area normalised H_{T}, for R15 reclustered jets for varying M_ϕ and r_{inv}.

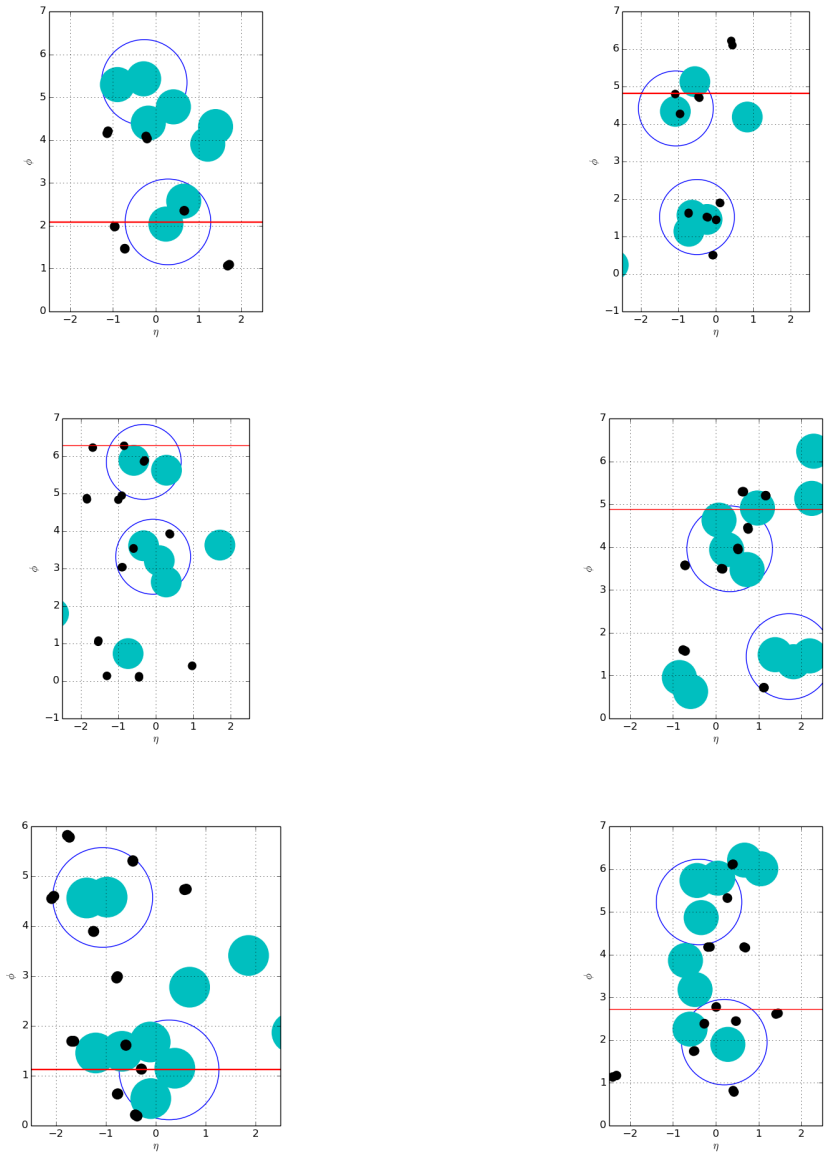


Figure A.2: Particle level objects plotted in the $\eta - \phi$ plane. The large hollow blue circles represent the reclustered jets, the filled cyan circles represent anti-k_t jets with R=0.4, the black points represent dark hadrons, and the red line the direction of missing transverses momentum.

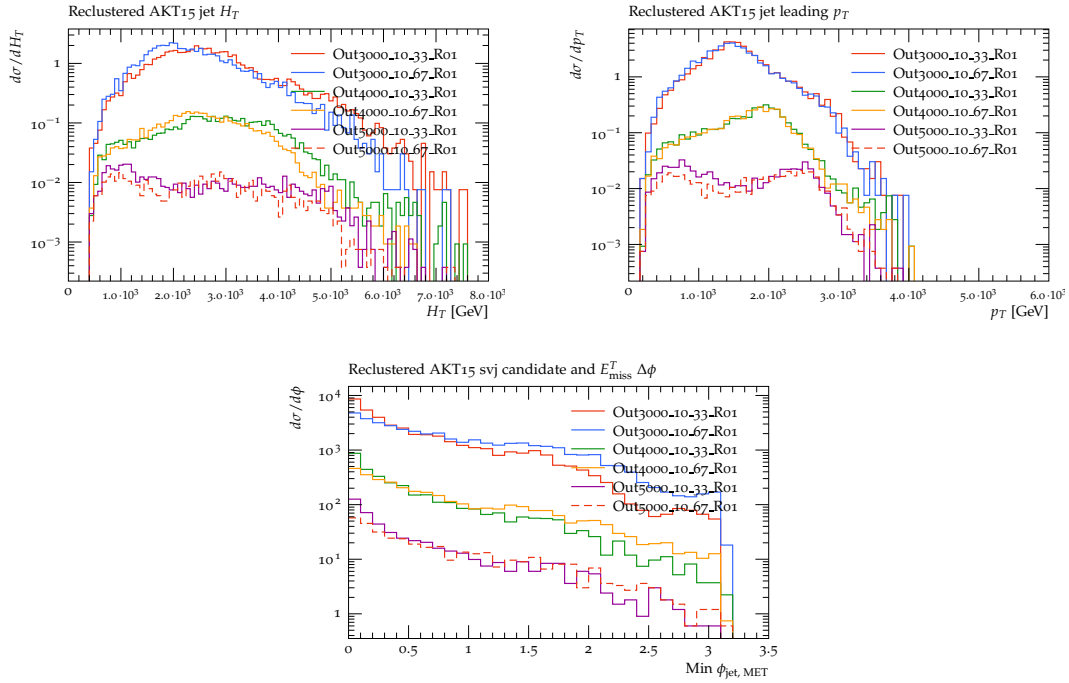


Figure A.3: Particle level distributions of H_{T}, leading jet p_{T} , and $\Delta\phi$(closest jet, $E_{\text{T}}^{\text{miss}}$) for R=1.0 jets for varying M_ϕ and r_{inv}.

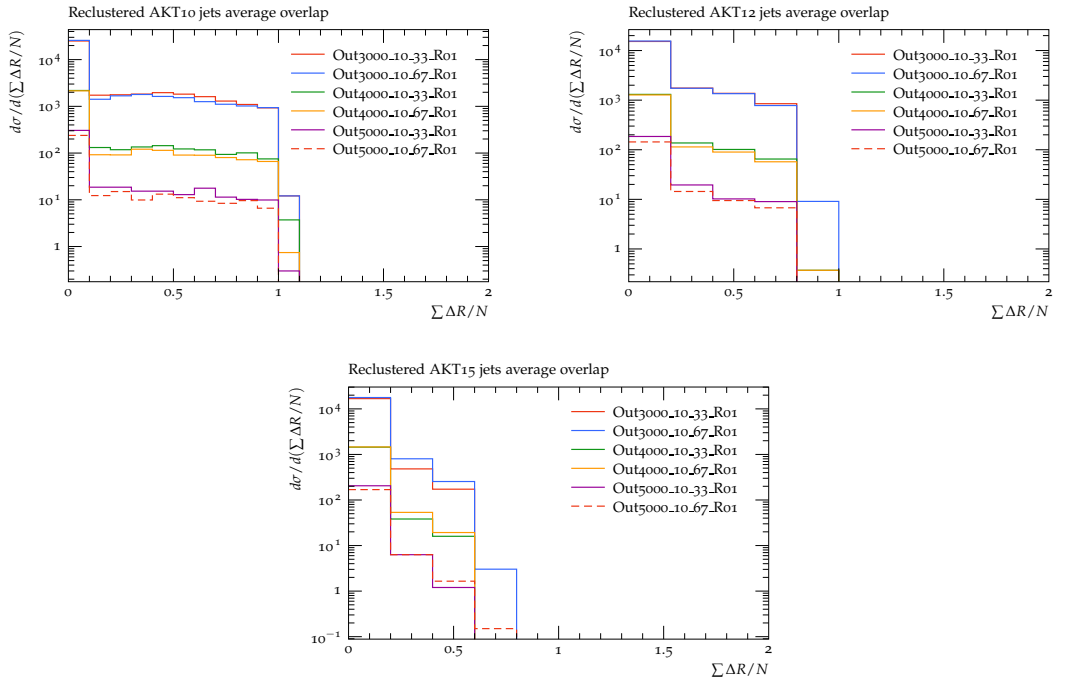


Figure A.4: Particle level distributions of average overlap between reclustered jets for R=1.0, R=1.2 and R=1.5 jets for varying M_ϕ and r_{inv}.

A.1.2 Detector level studies

Following the particle level discussion, reclustered jets were adapted as physics objects to be used in this analysis. The reclustering is performed by using anti-k_t R=0.4 jets as inputs.

As before, in Fig. A.5-A.6 show the objects in $\eta - \phi$ plane for six representative signal events, with r_{inv}= 0.6. Three different reclustered jet radii are considered: 1.0, 1.2, and 1.5. The black hollow circle denotes the SVJ candidate (having $E_{\text{T}}^{\text{miss}}$ closest to the jet), the blue hollow circle denotes jets satisfying the analysis threshold requirements, and the magenta hollow circle denotes jets which do not satisfy the analysis threshold requirements. The green filled circles represent anti-k_t R=0.4 jets with $p_{\text{T}} > 100$ GeV, the cyan filled circles represent anti-k_t R=0.4 jets with 50 GeV$< p_{\text{T}} < 100$ GeV, and the olive filled circles represent anti-k_t R=0.4 jets with $p_{\text{T}} < 50$ GeV.

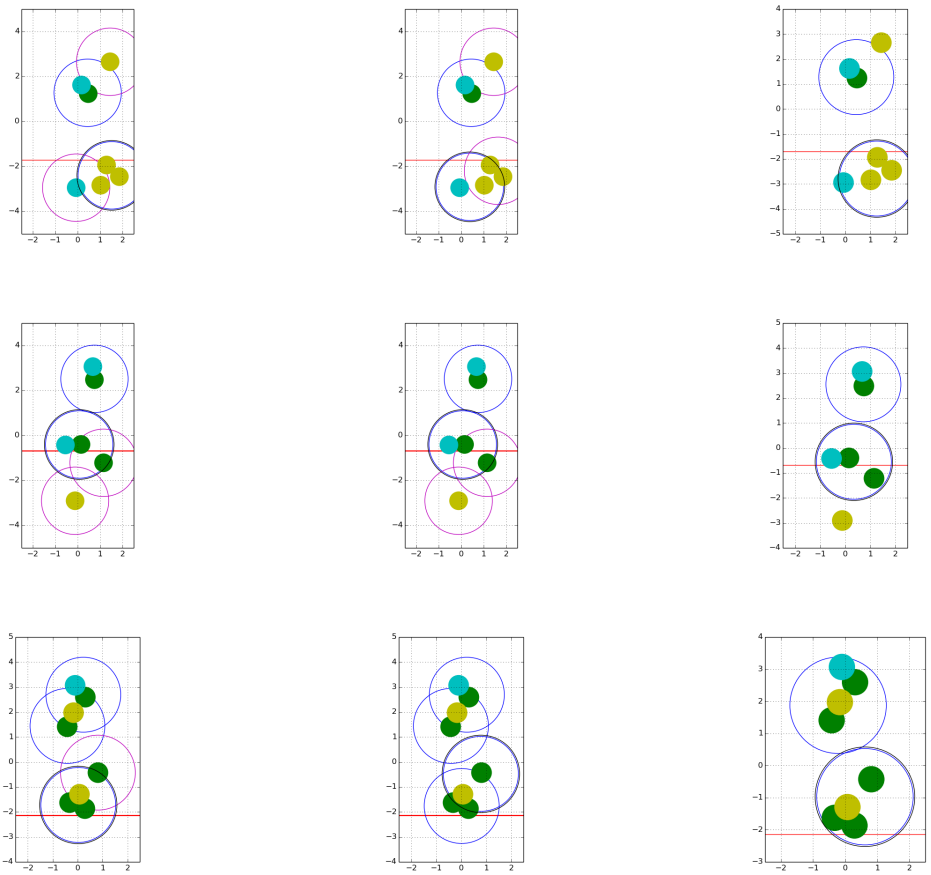


Figure A.5: Reconstruction level objects plotted in the $\eta - \phi$ plane for three events. Each row represents one event plotted for different radius parameters R = 1.0 (left), 1.2 (middle), 1.5 (right).

As can be seen from the above figures, selecting a larger radius value encompasses the semi-visible jets better and reduces the excess number of jets in the final state, similar to that seen in particle level. In all of the representative events, the $E_{\text{T}}^{\text{miss}}$ direction is within $\Delta\Phi < 1.0$ of the semi-visible jet.

Since the requirement of having jet mass > 40 GeV is a minimum necessity for proper jet performance, the choice of such R=1.5 jets, with leading jet $p_{\text{T}} > 300$ GeV is the most inclusive selection, that can be employed.

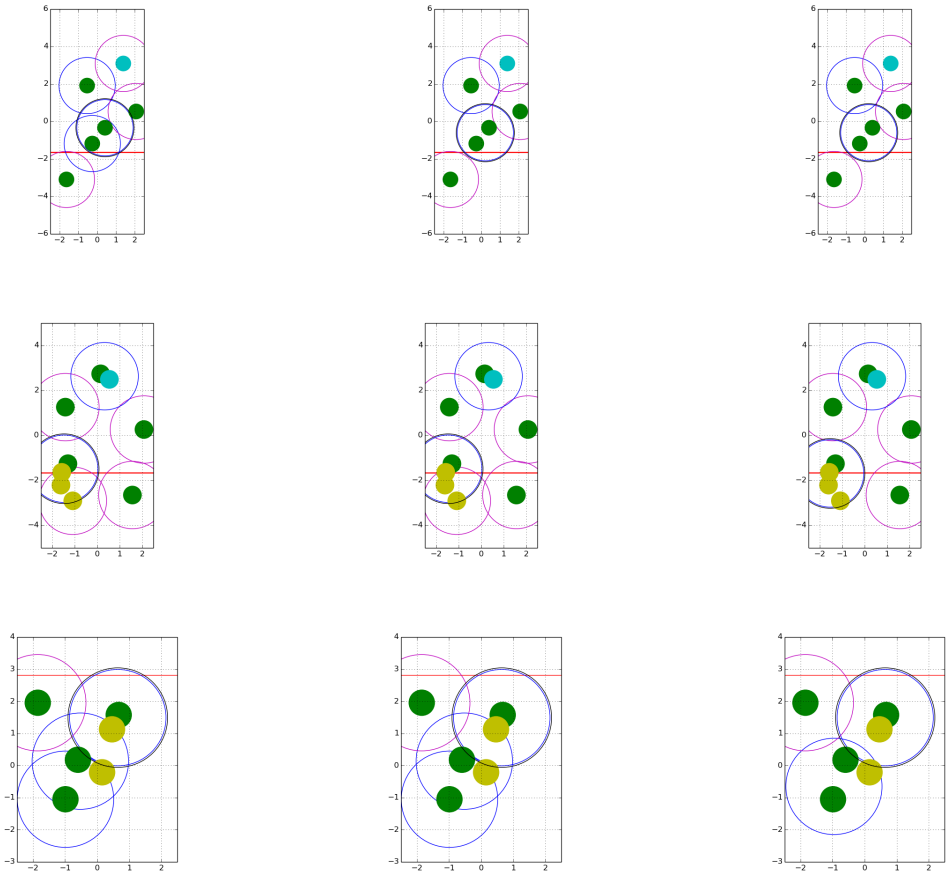


Figure A.6: Reconstruction level objects plotted in the $\eta - \phi$ plane for three events. Each row represents one event plotted for different radius parameters R = 1.0 (left), 1.2 (middle), 1.5 (right).

A.1.3 Performance

When moving to reclustered jets, it becomes necessary to check how differently these unusual jets (with dark hadrons) behave, when compared to standard q/g jets, if at all. Hence, a jet response study has been performed. The plots in Fig A.7 - A.10 shows the $p_{\mathrm{T}}^{rms}, p_{\mathrm{T}}^{mean}, \eta^{rms}$ and η^{mean} distributions for truth-matched reco-level anti-k_t R=0.4 jets. In most events, the reco-level jet is roughly in the same range as the corresponding particle-level jet, as can be seen from the η^{mean} and p_{T}^{mean} distributions, where the ratio is mostly close to unity.

However, it was observed that in most of the cases, only one subjet was seeding the reclustered jet, thereby leading to an unusual trend in the jet mass scale (JMS) distributions for low p_{T}/mass regimes, as can be seen from Fig A.11. Hence, reclustering the small radius jets to large radius jets is not a suitable option for the analysis.

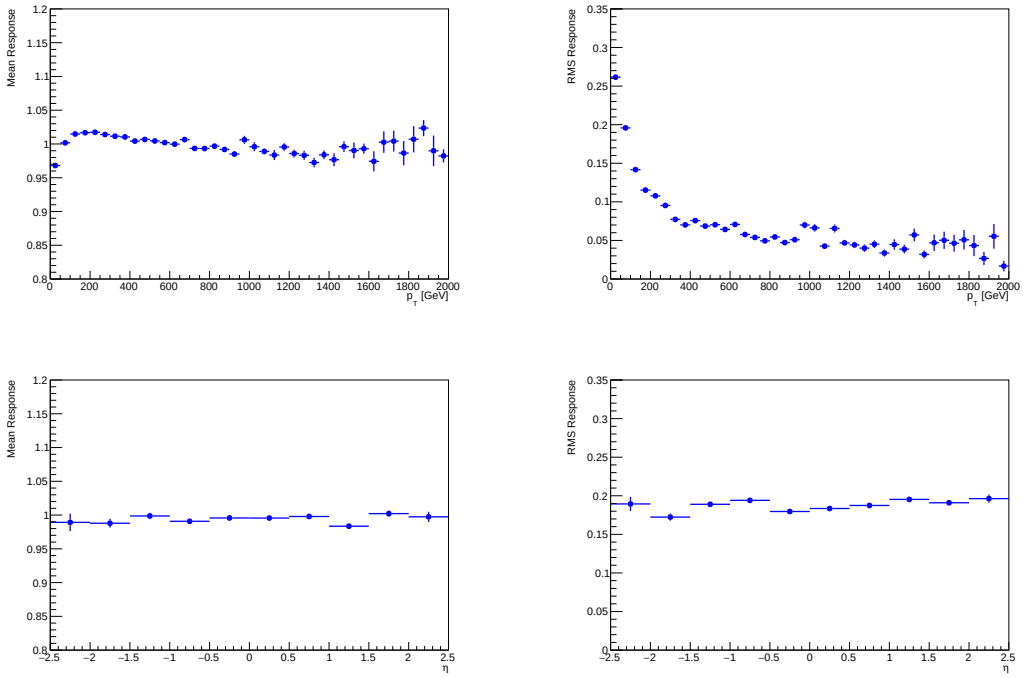


Figure A.7: $p_{\mathrm{T}}^{mean}, p_{\mathrm{T}}^{rms}, \eta^{mean}$ and η^{rms} distributions for truth-matched reco-level jets for signal point r_{inv}= 0.2.

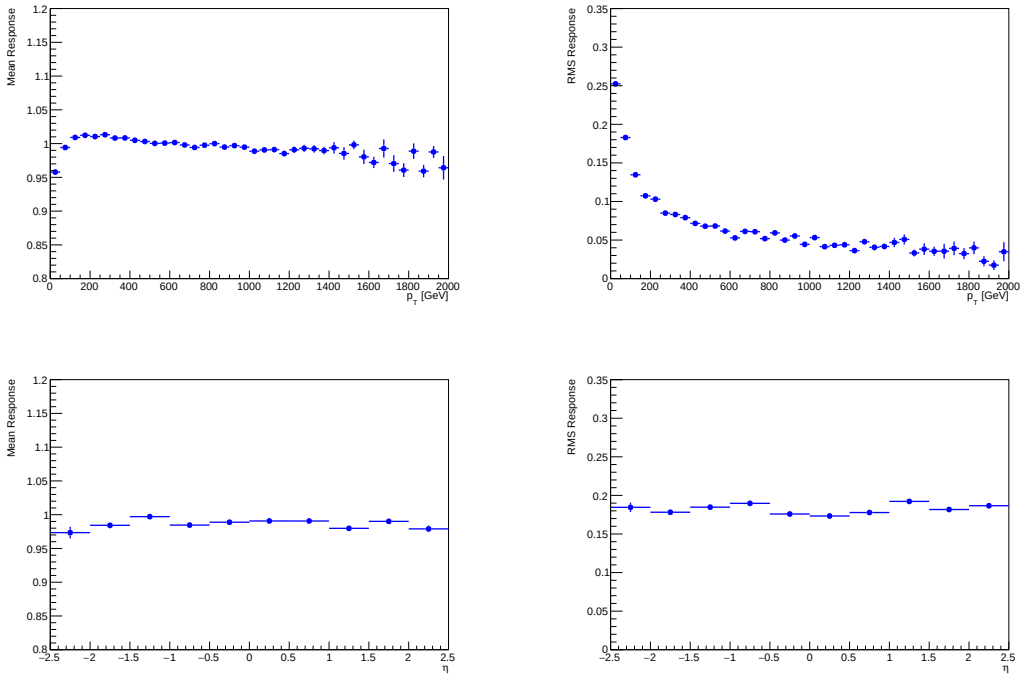


Figure A.8: $p_{\mathrm{T}}^{mean}, p_{\mathrm{T}}^{rms}, \eta^{mean}$ and η^{rms} distributions for truth-matched reco-level jets for signal point r_{inv}= 0.4.

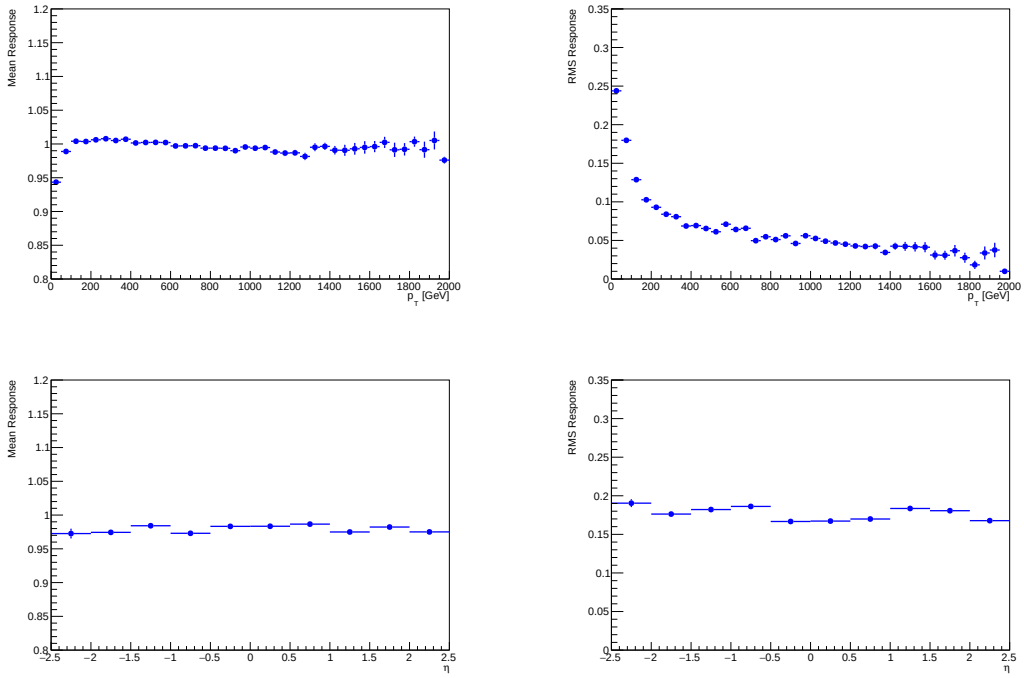


Figure A.9: $p_{\mathrm{T}}^{mean}, p_{\mathrm{T}}^{rms}, \eta^{mean}$ and η^{rms} distributions for truth-matched reco-level jets for signal point r_{inv}= 0.6.

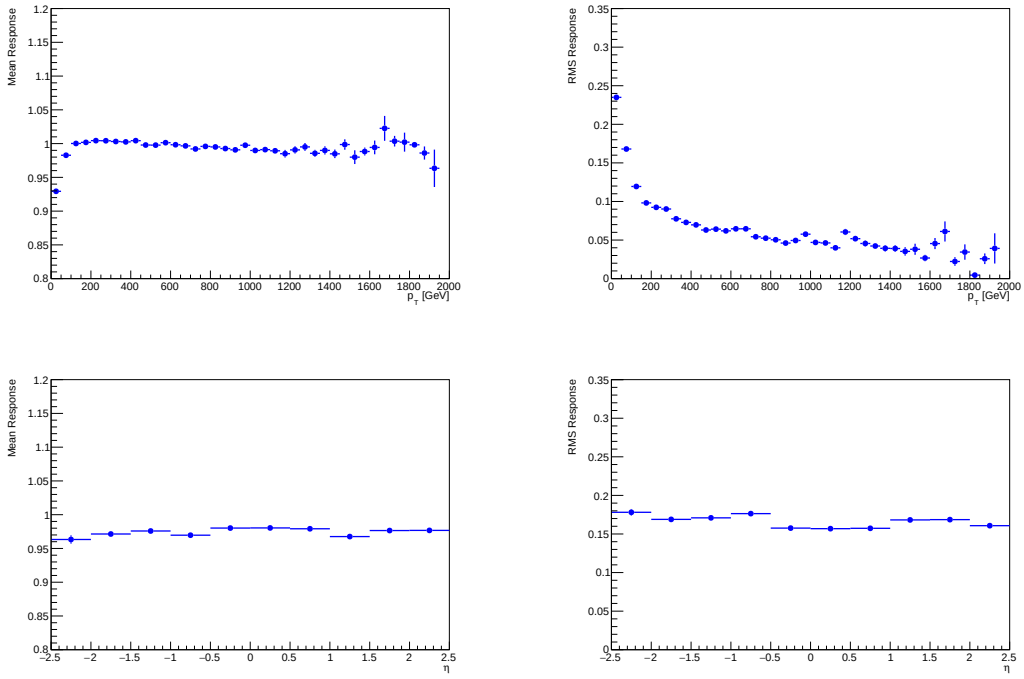


Figure A.10: $p_{\mathrm{T}}^{mean}, p_{\mathrm{T}}^{rms}, \eta^{mean}$ and η^{rms} distributions for truth-matched reco-level jets for signal point r_{inv}= 0.8.

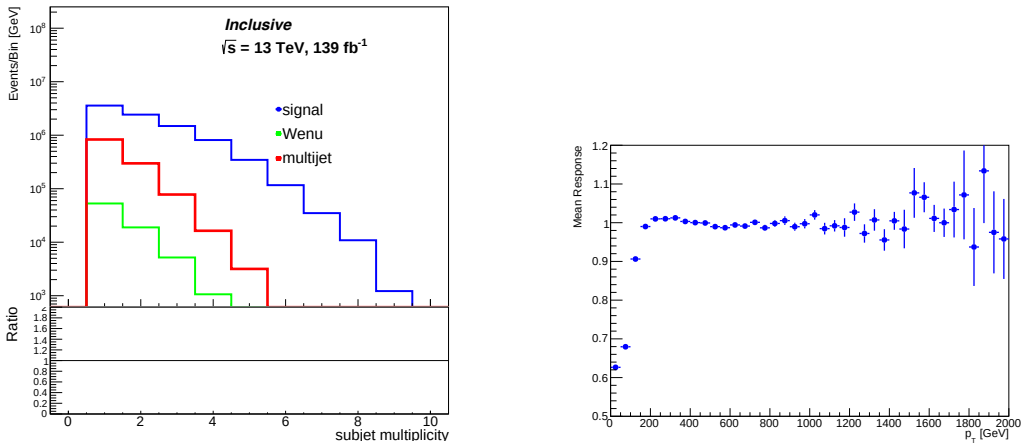


Figure A.11: Number of subjets seeding the reclustered jet (left). This leads to weird trend in JMS for low p_{T} and mass regime (right).

A.2 Tests for non-collisional background

To check for the impact of non-collisional background (NCB) on this analysis, leading jet ϕ was plotted for data, multi-jet MC and a representative signal mass point of 2.5 TeV, and $r_{\text{inv}}0.4$, as shown in Fig A.12. It was noticed that the shape agreement between data and MC fails at low leading jet ϕ tails. Hence, an additional jet selection was applied on data, i.e. the ratio of the jet charged-particle fraction to the maximum fraction of the jet energy collected by a single calorimeter layer, $f_{\text{ch}}/f_{\text{max}}$, is required to be larger than 0.1, on top of the existing overall jet selection criteria. The additional selection leads to a mild over correction as can be seen in the data vs. MC comparison plot. The same test was done on MC and signal separately, and the effect of NCB was found to be negligible in both cases.

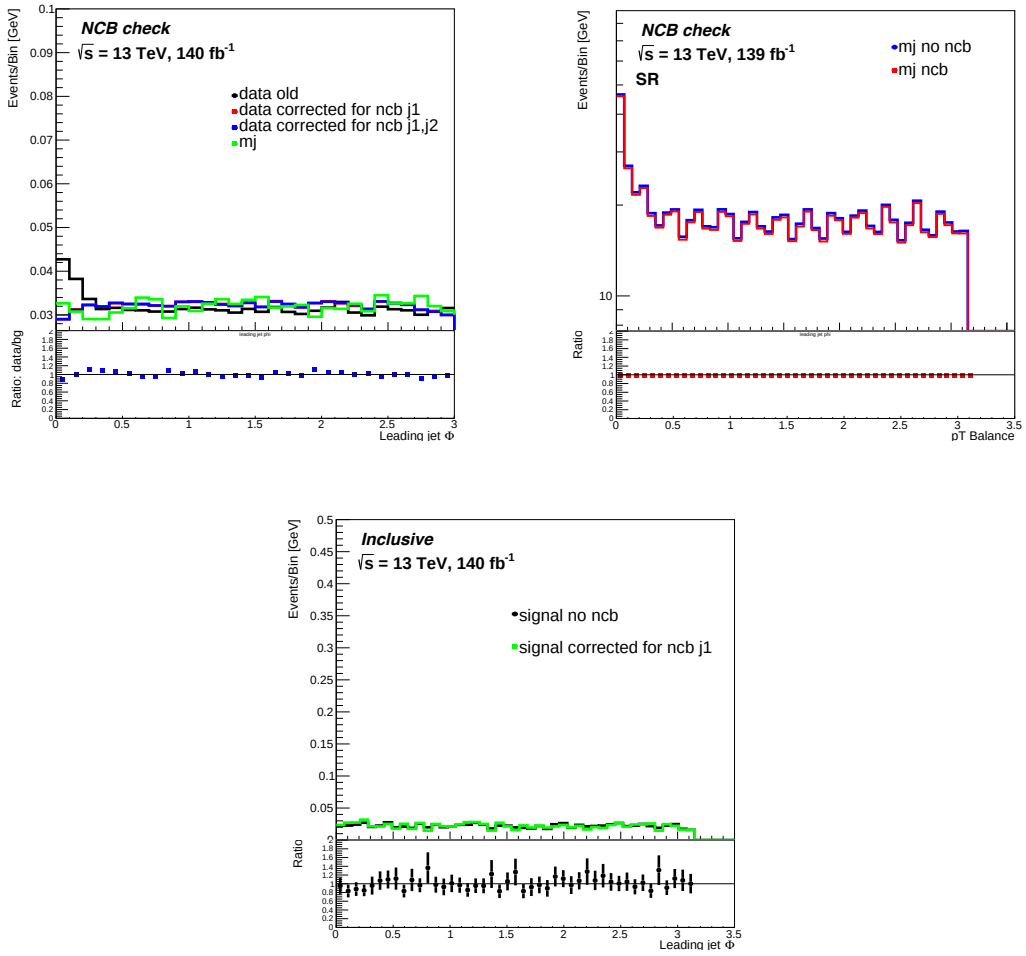


Figure A.12: Effect of NCB cleaning on data and MC.