CERN Summer Student Report 2017

Possible Improvements of the Signal Selections in ATLAS for Dark Matter Production in association with Heavy Flavour Quarks

Ophir Ruimi[1]
Supervisors: Priscilla Pani[2], Joaquin Poveda-Torres[2], Yoram Rozen[1]

[1]Technion – Israel Institute of Technology in Haifa
ophir.moshe.ruimi@cern.ch
rozen.yoram@gmail.com
[2]The European Organization for Nuclear Research (CERN),
Experimental Physics Department, Geneva, Switzerland
priscilla.pani@cern.ch
joaquin.poveda@cern.ch

CERN Summer Student Programme
September 2017

Contents

1. Introduction

During this Summer Student Programme at CERN I had the opportunity to join the DM+HF ATLAS group working on WIMPs pair production associated with heavy flavour (bottom and top) quarks, arising from proton-proton collisions at the LHC. My project focused on Dark Matter production involving bottom quarks, in $13.3\,\mathrm{fb}^{-1}$ of pp collisions data collected at a centre-of-mass energy of $\sqrt{s} = 13\ \mathrm{TeV}$ with the ATLAS detector at the LHC (Run 2) [1] [2]. I was using the ROOT analysis framework to discriminate background (reactions involving W and Z bosons jets, top-antitop quarks and top squarks, and others) from signal (those involving bottom-antibottom quarks).

1.1. The ATLAS experiment

Particle colliders represent one of the most useful tools currently available for conducting Particle Physics experiments and are therefore vital for improving our understanding of the universe. Particle colliders have facilitated the discovery of a plethora of new particles, as well as providing invaluable insights into the fundamental processes by which particles interact. With the recent discovery of the Higgs boson, made at the Large Hadron Collider (LHC) in 2012, it appears as if the Standard Model may be complete. However, there are still plenty of questions that need to be answered in the field of Particle Physics. In addition to in-depth study of the Higgs boson, there are many other directions for potential Particle Physics experiments, including: Top Physics studies; improved precision measurements of particle coefficients and masses; and the search for new Physics, beyond the Standard Model (BSM). [3] [4]

The LHC at CERN mentioned above, located near Geneva, Switzerland, is the world's largest particle collider. It accelerates and collides protons with unprecedented energies, allowing Physicists to test the predictions of different Particle Physics theories. ATLAS [5] (A Toroidal LHC ApparatuS, Fig. 1) is one of the seven particle detector experiments at the LHC, the largest in size. It is intended to investigate many different Physics phenomena that might become detectable in the energetic collisions of the LHC. Some of these are confirmations or improved measurements of the Standard Model, while many others are possible clues for new physical theories.

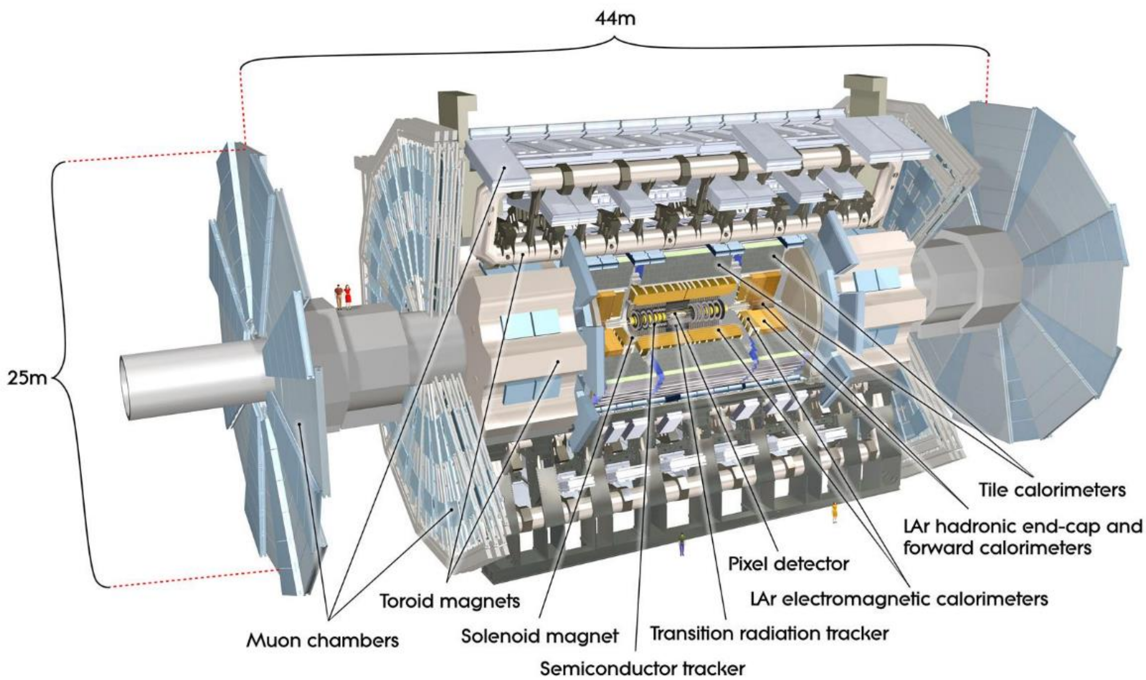


Figure 1: The ATLAS detector, image courtesy of ATLAS Experiment© 2012, CERN

The ATLAS detector is assembled in several layers around the nominal interaction point and symmetric in backward and forward direction along the beam pipe with respect to the interaction point. The detector consists of four main components: the inner detector measures the momentum of charged particles, the calorimeter measures the energies carried by the particles, the muon spectrometer identifies and measures the momenta of muons and the magnet system bends charged particles for momentum measurement. A right-handed coordinate system has its origin in the interaction point, the z-axis in the beam direction, the y-axis pointing upwards and the x-axis pointing towards the centre of the LHC ring; measuring these coordinates, it is easy to calculate the polar angles ϕ and θ. It is often useful [6] to define the pseudo-rapidity as:

$$\eta = -\ln\left[tan\left(\frac{\theta}{2}\right)\right]$$

and the distance ΔR in the $\eta - \phi$ space, that is:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

These variables, among others, define a coordinate system which is suitable for effitiently and comfortably describing and interpreting the experiments' results, and will be used in the following chapters.

1.2. DM production models

Astrophysical observations have provided compelling evidence for the existence of a non-baryonic dark component of the universe: Dark Matter (DM) [7] [8] [9] [10], which is not interacting with the electromagnetic force or emitting electromagnetic radiation such as light, and is thus invisible to the entire electromagnetic spectrum. The DM abundance has been precisely measured and corresponds to 26.8% of the mass-energy content of the universe [11] [12], while its nature remains largely unknown. One of the best motivated candidates for a DM particle is a Weakly Interacting Massive Particle (WIMP - χ) [13], which is assumed to be a Dirac fermion with a mass of 1 GeV. At the LHC, one can search for WIMPs produced in pp collisions. As they are weakly interacting, they escape the detectors undetected and are revealed by a momentum imbalance in the transverse plane of the event, the missing transverse momentum p_T^{miss} (and its magnitude E_T^{miss}).

Two choices for DM simplified models are considered in this analysis: one where the interaction with SM particles is mediated by a scalar (ϕ), and the second where only a light pseudo-scalar (a) is considered, assuming that the associated scalar is decoupled from the low-energy spectrum. Representative Feynman diagrams for tree-level production of these models is shown in Fig. 2.

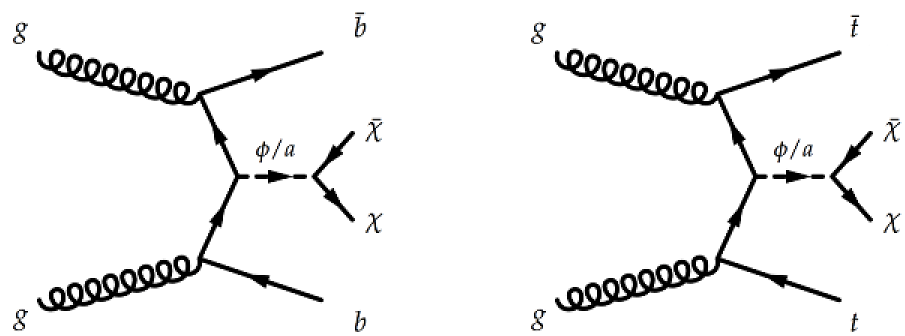


Figure 2: Representative Feynman diagrams at the lowest order for spin-0 mediator associated production with top and bottom quarks.

Lagrangian interaction terms for the analysis models:

$$L_\phi^{int} = -g_\chi \phi \bar{\chi}\chi - \sum_{fermions} g_q \frac{y_f}{\sqrt{2}} \phi \bar{f} f$$

$$L_a^{int} = -i g_\chi a \bar{\chi}\gamma^5 \chi - \sum_{fermions} i g_q \frac{y_f}{\sqrt{2}} a \bar{f}\gamma^5 f$$

2. Data analysis strategy

Particle Physics experiments require the careful analysis of large data samples, coming from an experimental apparatus, in order to measure the properties of fundamental particles. A very active field of research is focused on using these datasets to discover physical processes that have been predicted by theoretical models, but have not yet been observed in nature. Analyses generally rely on external predictions for the various background and signal components in the data to aid the interpretation of observations, where the signal component describes the process of interest. In Particle Physics, simulations of known and hypothesized Physics processes are run through a detailed detector simulation, and are subsequently reconstructed with the same algorithms as the data. In addition, background samples can be constructed using data-driven methods. The simulated samples may depend on one or many model parameters, for example the masses of hypothesized new particles such as foreseen by SUSY. It may be required, for instance when signals are analyzed over a multi-dimensional space of model parameters, to sample from a “grid” of potential signal scenarios, with each point on that grid corresponding to a unique point in the multi-dimensional parameter space. If no excess is observed in the data, exclusion limits may be set within this grid, excluding a subset of the tested parameter values. This analysis made use in the HistFitter framework [14] within the C++ based software ROOT developed in CERN.

Any Physics analysis aiming to study a specific phenomenon involves defining a region of phase space, obtained by applying selections to a set of kinematic observables, where a particular signal model predicts a significant excess of events over the predicted background level. Such a signal enriched region is called a Signal Region, or SR. To estimate background processes contaminating the SR(s) in a semi-data-driven way, one typically defines Control Region(s), or CR(s), in which the dominant background(s) can be controlled by comparison to the data samples. CRs are specifically designed to have a high purity for one type of background, and should be free of signal contamination. A third important component of data analysis is the validation of the model used to predict the number of background events in the SR(s). Validation Region(s), VR(s), are defined for this purpose. VR(s) are typically placed in between the CR(s) and SR(s). Hence, the choice of VR(s) is typically a trade-off between maximizing statistical significance and minimizing signal contamination, while controlling the assumptions in the extrapolation from CR(s) to SR(s). The concepts of CRs, VRs and SRs and the extrapolation between them are schematically shown in Figs. 3 and 4.

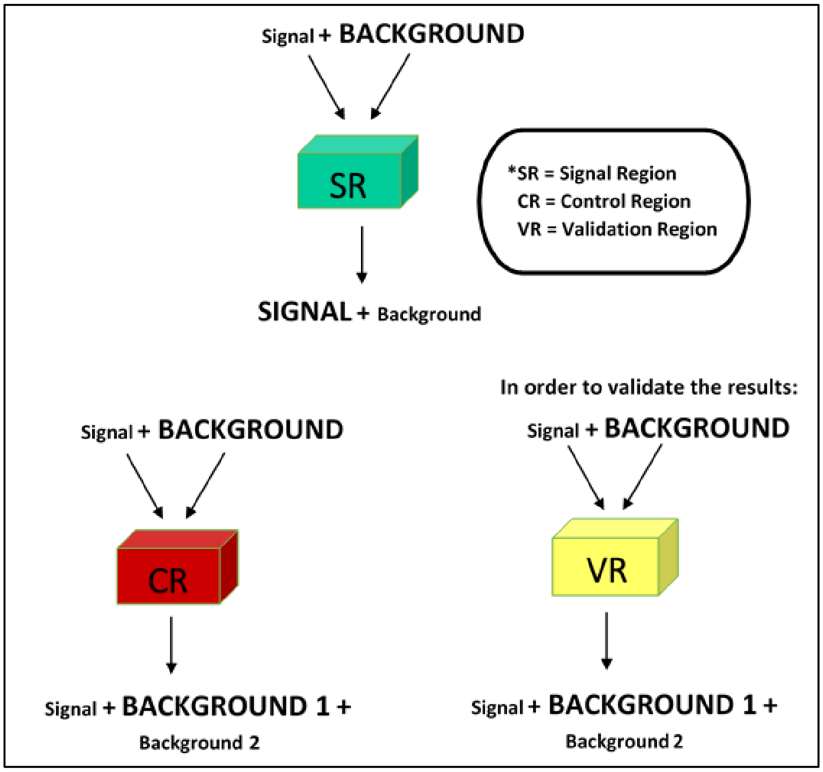


Figure 3: A schematic view of the use of the signal, control and validation regions.

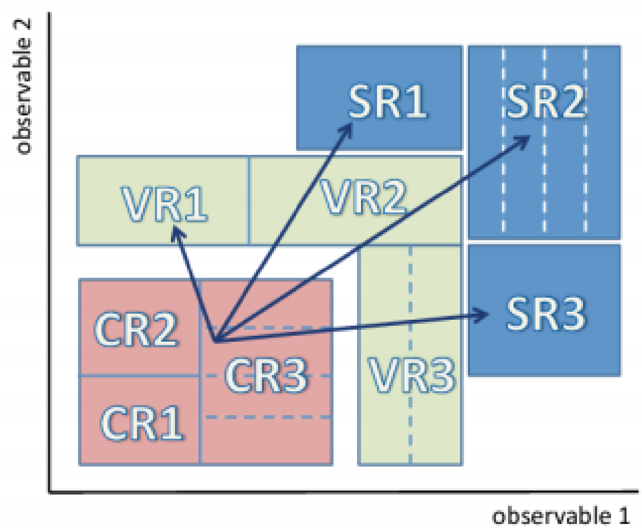


Figure 4: [14] A scheme of an analysis strategy with multiple signal, control and validation regions. All regions can have single or multiple bins, as illustrated by the dashed lines. The extrapolation from the control to the signal regions is verified in the validation regions that lie in the extrapolation phase space, and shown by the arrows on the figure.

3. Event selection and background estimation

Data used in this analysis were collected by the ATLAS detector in pp collisions at the LHC using a centre-of-mass energy of 13 TeV and 25 ns proton bunch crossing interval during 2015 and 2016. After requiring beam, data and detector-quality criteria, the available dataset corresponds to an integrated luminosity of 13.3 fb^{-1} with an associated uncertainty of ±4%. The search for dark matter produced in association with bottom quarks is based on a selection of events with b-jets and missing transverse momentum in the final state. Events containing leptons (electrons or muons, as taus are not relevant due to their very different behaviour) are explicitly vetoed in the signal regions, and are used to define control and validation regions. Events are required to pass a trigger selection based on E_T^{miss}. An overlap removal procedure is applied to prevent double-counting of reconstructed objects. Jets are b-tagged by a multivariate algorithm which assigns a "weight" to each jet using information on the impact parameters of inner detector tracks associated to the jet, the presence of displaced secondary vertices, and the reconstructed flight paths of b- and c-hadrons inside the jet. The missing transverse momentum p_T^{miss} is defined as the negative vector sum of the p_T of all selected and calibrated physical objects in the event, with an extra term added to account for soft energy in the event that is not associated to any of the selected objects. This soft term is calculated from inner detector tracks with p_T > 0.4 GeV matched to the primary vertex to make it more resilient to pileup contamination.

After measurement, the data, as C++ objects, were stored in files within ROOT's tree data structure (leaves, branches, trees and so forth), to be extracted, reconstructed and analyzed using a multivariate analysis method. For the analysis, we used, among others, the following basic ROOT classes: TFile (in order to get the data from the data source); TTree (that stores arrays of variable entries and fills them with data); TBrowser (to interactively inspect ROOT files); the histogram classes derived from TH1 (TH1D, TH1F, etc.) that handle frequency distributions (e.g., in TH1F, "H" stands for "histogram", "1" for 1D i.e. a one variable graph and "F" for "float" – meaning that the data type *float* is used to store the entries in a single histogram bin); TCanvas (creates a canvas for the graphical output file); THStack (allows to manipulate a set of histograms as a single entity), and others.

This analysis is targeting signals with an event topology characterised by two b-jets and missing transverse momentum originated from the invisible decay of a scalar or pseudoscalar mediator to the dark sector. Therefore, only events with E_T^{miss} > 150 GeV and exactly two b-tagged jets are considered in the signal selections.

The signal is characterised by low jet multiplicity, therefore the contamination from high jet multiplicity backgrounds, particularly $t\bar{t}$ production, is suppressed by vetoing events containing a third jet with p_T > 60 GeV. The events in the SR are required to have all jets characterised by a large angular separation: ΔR_{min} = min{ΔR_{ij}} > 2.8, i≠j ϵ {1,2,3}. A requirement on the separation in pseudorapidity and azimuthal separation between the two b-jets is required as *Δη(b1, b2)* > 0.5 and *Δϕ(b1, b2)* > 2.2 rad (the difference between the azimutal distances between E_T^{miss} and each of the two jets in the event). Finally, to further enhance the signal yields over the SM background, the events are required to have a transverse momentum imbalance between the two b-jets: $Imb(b1, b2) = \frac{p_T(b_1) - p_T(b_2)}{p_T(b_1) + p_T(b_2)}$ > 0.5. The requirements for all regions are summarised in table 5 below.

Quantity	SR	CRZ1b	VRZ2b	CRW1b	VRW1b	CRW2b	VRLR
N_{lepton} (baseline)	0	2 (SFOS)	2 (SFOS)	1	1	1	0
N_{lepton} (high-purity)	0	2 (SFOS)	2 (SFOS)	1	1	1	0
$\Delta\phi^j_{\text{min}}$	> 0.4	> 0.4	> 0.4	> 0.4	> 0.4	> 0.4	> 0.4
N_{jets}	$2-3$	$2-3$	$2-3$	$2-3$	$2-3$	$2-3$	$2-3$
$N_{\text{b}jets}$	$= 2$	$= 1$	$= 2$	$= 1$	$= 1$	$= 2$	$= 2$
jet 1 p_{T} [GeV]	> 100	> 100	> 85	> 100	> 100	> 100	> 100
jet 2 p_{T} [GeV]	> 20	> 20	> 20	> 30	> 30	> 20	> 20
jet 3 p_{T} [GeV]	< 60	< 60	< 60	< 60	< 60	< 60	< 60
$p_{\text{T}}^{\text{b-jet1}}$ [GeV]	> 50	> 50	> 50	> 50	> 50	> 50	> 50
$E_{\text{T}}^{\text{miss}}$ [GeV]	> 150	< 100	< 80	> 130	> 150	> 120	> 150
$E_{\text{T}}^{\text{miss,cor}}$ [GeV]	-	> 120	> 100	-	-	-	-
ΔR_{min}	> 2.8	> 2.8	> 2.8	> 2.5	> 2.8	> 2.8	< 2.5
$\Delta\eta(b_1, b_2)$	> 0.5	-	-	-	> 0.5	-	> 0.5
$Imb(b_1, b_2)$	> 0.5	-	-	-	-	-	> 0.5
m_{T}^{lep}	-	-	-	[30, 100]	[30, 100]	> 30	-
$m_{\ell\ell}$	-	[75, 105]	[80, 100]	-	-	-	-
lepton 1 p_{T} [GeV]	-	> 30	> 30	> 30	> 30	> 30	-
lepton 2 p_{T} [GeV]	-	> 25	> 25	-	-	-	-
$\Delta\phi(b_1, b_2)$	> 2.2	> 2.2	-	[1, 2.2]	> 2.2	> 2.2	> 2.2

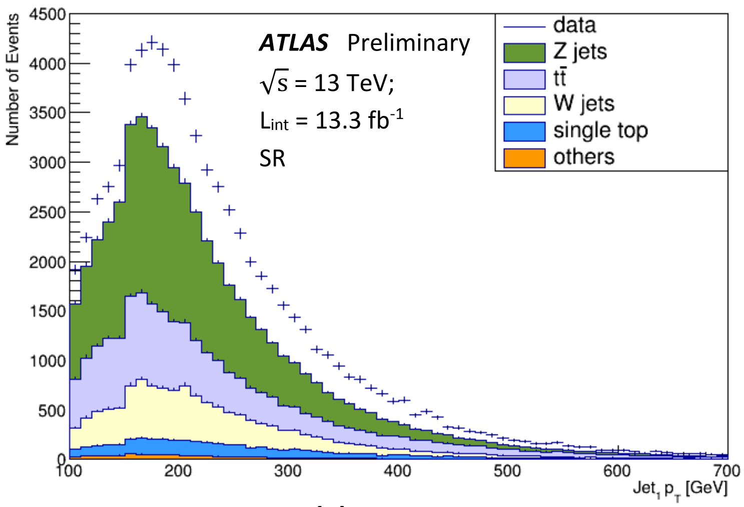
Table 5: [1] Summary of the selections of the signal, control and validation regions of the analysis.

For a signal model corresponding to $m_{\Phi(a)}$ = 20 GeV and m_χ = 1 GeV, approximately 3×10^{-6} of the simulated signal events are retained by the SR selections for both scalar and pseudo-scalar mediators. The dominant SM background process is the production of Z bosons in association with heavy-flavour jets. This is followed by the single and pair production of top-quarks. Top-quark pair production contributes to the signal region when a charged lepton is produced but the event is not rejected, either because the lepton is a hadronically decaying τ, or because the electron or muon is not identified or out of detector acceptance. Single-top production is a subdominant contribution in the signal region and is almost entirely due to t-channel production.

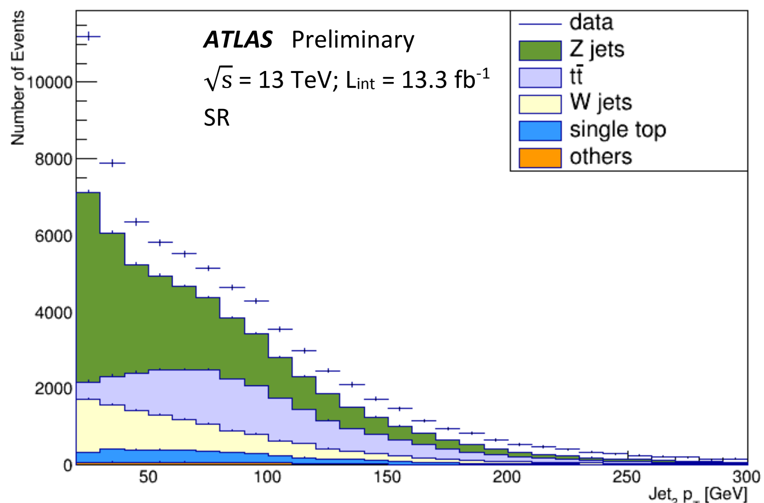
4. Results

This analysis included measurements triggered within the signal regions defined above. A characteristic distribution for the SRs is shown in figure 6.

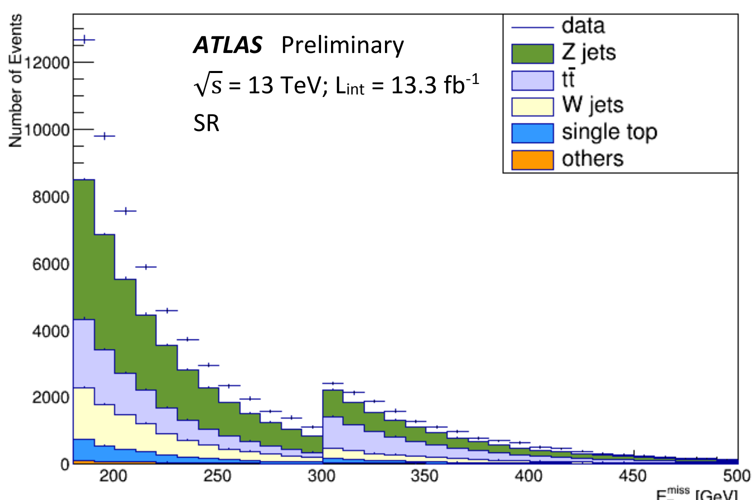
The branches *Δη(b₁, b₂)*, *Δϕ(b₁, b₂)* and $E_T^{miss, cor}$ (lepton corrected) were not contained in the signal region and yielded no data. Histograms for the rest of the parameters in the CRs and VRs, as well as ΔR_{min} and *Imb*(b_1,b_2) in the SR, are found in [1].



(a) SR, jet 1 p_T



(b) SR, jet 2 p_T



(c) SR, E_T^{miss}

Figure 6: Representative distributions for the signal region of the analysis. All backgrounds are normalised to the fit results.

As shown in [1] (ch. 6), in the absence of a sign of an excess of events over expected backgrounds, the results are translated into 95% confidence level (CL) upper limits on contributions from BSM Physics for the signal region, assuming no systematic uncertainties for these events and neglecting any possible contamination in the control regions.

5. Summary

The result of a search for dark matter pair production in association with bottom quarks in 13.3 fb^{-1} of pp collisions collected at $\sqrt{s}$ = 13 TeV by the ATLAS detector at the LHC was presented. The search was optimised for spin-0 mediators to the dark sector which are produced in association with bottom quarks and decay into a pair of Dirac Dark Matter particles. Events containing large missing transverse momentum and exactly two jets identified as originating from b-quarks were considered. The data were found to be consistent with the SM background expectations and the results were interpreted in terms of 95% CL exclusion limits on the production cross section of a spin-0 mediator in association with b-quarks as a function of the mediator mass, assuming the DM particle mass to be m_χ = 1 GeV.

6. Acknowledgement

I was honoured to participate the Summer Student Programme which is a unique opportunity to join the ongoing work at the forefront of groundbreaking HEP research conducted at the LHC, to gain practical knowledge and to establish long-term relationships and cross-fertilization with students and scientists across the globe at the “melting pot” called CERN. This has been an enriching once-in-a-lifetime experience and I am certain I will return to CERN as an integral part of my academic future.

I would like to thank my CERN supervisors, Pricilla Pani and Ximo Poveda-Torres, and my project supervisor from the Technion, Yoram Rozen, for their careful guidance, patience and dedication. My great gratitude is given to Vincent Cecchini, Kevissen Sellapillay and Tigran Mkrtchyan for their daily software assistance. I thank Shlomit Tarem, Yael Shadmi, Revital Kopeliansky, Yevgeny Kats, Yoav Afik, Eitan Gozani and Yaniv Weiss for useful discussions, and Abraham Berman, David Sharoni, Yuval Cohen, Ilan Rusnak and Francois Bodenheimer for their significant contribution.

7. References

[1] ATLAS-CONF-2016-086 “Search for Dark Matter production associated with bottom quarks in 13.3 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector at the LHC”, 8 August 2016.

[2] “Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum”, arXiv: 1507.00966v1 [hep-ex], 3 Jul 2015.

[3] ATLAS Collaboration: Aad, G., et al., “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, Physics Letters B, Volume 716, Issue 1, Pages 1-29, ISSN 0370-2693, http://dx.doi.org/10.1016/j.physletb.2012.08.020, 17 September 2012.

[4] CMS Collaboration: Chatrchyan, S., et al., “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, Physics Letters B, Volume 716, Issue 1, Pages 30-61, ISSN 0370-2693, http://dx.doi.org/10.1016/j.physletb.2012.08.021, 17 September 2012.

[5] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, JINST 3:208003, (2008).

[6] A. Bettini, “Introduction to Elementary Particle Physics”, 2nd Ed. (Cambridge University Press, 2014).

[7] F. Zwicky, “Die Rotverschiebung von extragalaktischen Nebeln”, Helv. Phys. Acta 6 (1933) 110–127.

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

[9] G. Jungman, M. Kamionkowski and K. Griest, Supersymmetric dark matter, Phys. Rept. 267 (1996) 195–373, arXiv: hep-ph/9506380 [hep-ph].

[10] J. Binney and S. Tremaine, ’Galactic Dynamics’, 1993, isbn: ISBN:9781400828722.

[11] WMAP Collaboration, Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation, Astrophys. J. Suppl. 192 (2011) 18, arXiv:1001.4538 [astro-ph.CO].

[12] P. A. R. Ade et al., Planck 2015 results. XIII. Cosmological parameters, 2015, arXiv: 1502.01589 [astro-ph.CO].

[13] G. Steigman and M. S. Turner, “Cosmological Constraints on the Properties of Weakly Interacting Massive Particles”, Nucl. Phys. B 253 (1985) 375.

[14] “HistFitter software framework for statistical data analysis “, arXiv: 1410.1280v1 [hep-ex], 6 Oct 2014.