
Tracker Readout Development
for the CMS Upgrade
and a
Search for Light Higgs Bosons
from SUSY Decay Cascades

by

MEI-LI HOLMBERG



SCHOOL OF PHYSICS
UNIVERSITY OF BRISTOL

JULY 2023

Abstract

This thesis focuses on two areas of the CMS experiment at the Large Hadron Collider (LHC): the upgrade work for the CMS detector at the High-Luminosity LHC (HL-LHC) and the Run 3 preparations for a Next-to-Minimal Supersymmetric Standard Model (NMSSM) search for light Higgs bosons.

The HL-LHC will provide a higher instantaneous luminosity than the current LHC, resulting in a larger volume of data for physics analysis. However, this demands improvements in the triggering systems. In response, CMS is developing an innovative Level-1 (L1) hardware track trigger using data from the Outer Tracker. This thesis discusses the L1 Track Finder, which employs a pattern recognition algorithm implemented on Field Programmable Gate Arrays (FPGAs). The L1 Track Finder's primary objective is to reconstruct within $4\text{ }\mu\text{s}$ the trajectories of all charged particles detected within the Outer Tracker volume with transverse momenta greater than 2 GeV . The CMS upgrade work also includes testing of high-speed optical transceivers, which are crucial for efficient data transmission between the CMS backend systems. These transceivers are required to transmit data at a rate of 25 Gb/s with good signal quality. Performance studies of various optical transceivers are presented.

Furthermore, the thesis discusses a search for light Higgs boson decays produced in SUSY decay cascades, in the context of the NMSSM. The analysis looks for events with little missing transverse momentum, where the light Higgs bosons decay to boosted b quark pairs. The analysis builds upon a previously published study conducted using Run 2 data. The current work explores improvements for the upcoming Run 3 data from changing the double- b -tag discriminator and jet mass reconstruction to using the ParticleNet algorithm.

Declaration

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific references in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: DATE:

Author Contributions

Chapters 1-2 provide a general overview of particle physics, the Large Hadron Collider (LHC), and the CMS detector, and do not contain any original work.

In Chapter 3, I contributed to the development of the L1 tracking algorithm firmware as part of a larger group. Specifically, I was in charge of creating, testing and maintaining the HLS firmware for the VMRouter processing module and the VMRouter for Combined Modules, which are discussed in Section 3.5 and Section 3.7.1, respectively. My project efforts also consisted of testing and updating the firmware, including the intermediate memories and other processing modules described in Sections 3.4.1-3.4.2. Additionally, I collaborated in developing the scripts used to generate the top-level VHDL for track-finding chains (Section 3.4.3), and I added to the scripts the ability to create VHDL testbenches for the generated chains. I also created and updated test utilities to import emulated test data for use as input to the firmware simulations and for comparing the firmware output to the emulation output. These scripts and utilities were then used to simulate the “Skinny” chain and the “Barrel-Only” chain discussed in Section 3.6, where I was personally involved in validating the former.

The study on the missing transverse momentum trigger efficiency, presented in Chapter 4, was performed by me using pre-existing emulations and plotting tools.

Chapter 5 details the group effort to test optical transceivers for the Serenity back-end board. Within the group, I updated test scripts for the parameter scanning (Section 5.4), the optical oscilloscope measurements (Section 5.5), and the attenuation scanning (Section 5.6). I conducted a large part of the testing, however, the data in Section 5.4.3 and Appendix B were taken by Alex Howard. The PRBS comparisons in Section 5.5.1 were performed by Giacomo Fedi. Additionally, I wrote most of the plotting scripts for the results and implemented the receiver measurements in Section 5.7.

Based on the common CMS data analysis tool, NANOAODTOOLS, I developed the majority of the analysis framework and plotting scripts for the NMSSM search for light Higgs bosons in Chapter 6. As part of this effort, I conducted comparisons between the btagHbb and ParticleNet double-b-tag discriminators (Sections 6.7 and 6.10) using Monte Carlo samples created by Jacob Linacre and Giovanna Salvi. I also implemented and studied the QCD background estimation (Section 6.8) and some of the systematic uncertainties (Section 6.9). Additionally, I have made contributions to the NANOAODTOOLS software itself.

Acknowledgements

I would like to thank my supervisors, Joel Goldstein and Ian Tomalin (shown in Figure 0), for their guidance and support throughout the course of my doctoral research. Their expertise and their ability to point out all my silly mistakes have shaped the outcome of this thesis.



Figure 0: A very important meeting with the author (left) and her supervisors Joel Goldstein (middle) and Ian Tomalin (right).

I am also grateful to Emyr Clement for his expertise in studying the trigger efficiencies, as well as to Alex Howard and Greg Iles for their assistance in studying the optical transceivers for Serenity. Furthermore, I would like to thank Jacob Linacre for his guidance in the search for light Higgs bosons.

Lastly, I would like to express my gratitude to my family, friends, and my dog for distracting me from my work.

Contents

List of Figures	viii
List of Tables	xiii
1 Background	1
1.1 Introduction	1
1.2 The Standard Model	1
1.2.1 Fermions	2
1.2.2 Bosons	4
1.2.3 Physics Beyond the Standard Model	5
1.3 Supersymmetry	6
1.3.1 Supersymmetric Particles	7
1.3.2 Next-to-Minimal Supersymmetric Standard Model	9
2 The LHC and the CMS Detector	11
2.1 Introduction	11
2.2 The Large Hadron Collider	12
2.3 The CMS Detector	15
2.3.1 Solenoid Magnet	16
2.3.2 Tracker	17
2.3.3 Electromagnetic Calorimeter	19
2.3.4 Hadron Calorimeter	20
2.3.5 Muon Detectors	20
2.3.6 Trigger System	21
2.3.7 Offline Event Reconstruction	23
2.4 High-Luminosity LHC	26
2.5 The CMS Phase-2 Upgrade	26
2.5.1 Outer Tracker	29
3 L1 Tracking Algorithm	35
3.1 Introduction	35
3.2 Track-Finding Algorithm	36
3.2.1 Virtual Modules	37
3.3 Field Programmable Gate Arrays	39

3.3.1	Vivado and Vivado HLS	40
3.4	Track-Finding Firmware Implementation	41
3.4.1	Processing Modules	43
3.4.2	Intermediate Memories	47
3.4.3	Top-Level VHDL	47
3.5	VMRouter	48
3.5.1	Resource Usage and Timing	50
3.6	Simulations and Demonstrations	53
3.6.1	Skinny Chain	53
3.6.2	Barrel-Only Chain	54
3.7	Combined Modules	55
3.7.1	VMRouter for Combined Modules	56
3.8	Conclusions	58
4	Study of the Missing Transverse Momentum Trigger Efficiency	60
4.1	Introduction	60
4.2	Method	61
4.3	Results	62
4.3.1	Background Trigger Rates	62
4.3.2	Missing Transverse Momentum Trigger Efficiency	64
4.3.3	Jet p_T and H_T Trigger Efficiencies	64
4.4	Conclusions	67
5	Study of Optical Transceivers for Serenity	68
5.1	Introduction	68
5.2	Optical Transceivers	69
5.2.1	Firefly 12-Channel Transceiver	69
5.2.2	Firefly 4+4-Channel Transceiver	70
5.3	Test Setup	71
5.4	Parameter Scanning	74
5.4.1	FPGA Transceiver Parameters	75
5.4.2	Firefly Parameters	76
5.4.3	Results	76
5.5	Optical Oscilloscope Measurements	83
5.5.1	Results	85
5.6	Attenuation Scanning	90
5.6.1	Results	91
5.7	Receiver Measurements	95
5.7.1	Results	97
5.8	Conclusions	99
6	Search for Light Higgs Bosons from SUSY Decay Cascades	101
6.1	Introduction	101

6.2	Double- b -tagging	102
6.2.1	btagHbb	103
6.2.2	ParticleNet	103
6.3	Jet Mass Reconstruction	104
6.3.1	Softdrop Mass	104
6.3.2	ParticleNet	105
6.4	Signal Model	105
6.5	Event simulation	107
6.6	NanoAODTools	108
6.7	Event Selection	108
6.7.1	Kinematic Event Selection	109
6.7.2	Double- b -tag Selection	111
6.7.3	Mass Event Selection	116
6.7.4	Signal Mass Region	126
6.8	Background Estimations	127
6.8.1	QCD Data-driven Background Estimation	127
6.9	Systematic Uncertainties	129
6.9.1	Double- b -tag Scale Factor	131
6.9.2	Jet Mass Systematic Uncertainties	132
6.10	ROC Curves	133
6.11	Improvements	136
6.11.1	Double- b -tag cuts	136
6.11.2	Mass Region Bins	137
6.11.3	Data-driven Background Estimation for $t\bar{t}$ Background .	137
6.12	Conclusions	138
7	Conclusions	140
	Appendix A Firefly Alpha Transceiver Measurements	143
	Appendix B Firefly Beta Transceiver Measurements on Serenity	
	HNZN	147
B.1	Parameter Scan	147
B.1.1	Clock and Data Recovery Issue	150
	Appendix C Additional NMSSM Results	152
C.1	Mass Distribution of Signal Samples	152
	Acronyms	158
	Bibliography	159

List of Figures

1.1	All elementary particles in the Standard Model.	2
1.2	Examples of Feynman diagrams of quantum loop corrections for the Higgs boson mass.	5
1.3	A triangle loop Feynman diagram of three gauge bosons connected to a fermion.	8
2.1	A simplified illustration of the data-flow between the CMS detector and the back-end systems.	12
2.2	The LHC and the locations of the four main detectors.	13
2.3	The integrated and instantaneous luminosity delivered at the CMS experiment.	14
2.4	The CMS detector and the location of its subsystems.	16
2.5	A slice of the CMS detector and how different types of particles interact with the various subdetectors.	17
2.6	The Phase-1 tracker layout.	18
2.7	An illustration of jets.	25
2.8	An illustration of boosted double- b jets.	25
2.9	A timeline of the LHC and the HL-LHC projects.	27
2.10	The tracker layout for the HL-LHC.	28
2.11	The stub-finding concept, and a depiction of an Outer Tracker sensor module.	31
2.12	A diagram of the data-flow and latency requirements from the Outer Tracker front-end modules to a Level-1 Accept via the back-end electronics.	32
2.13	The detector and track-finding sectors in an $x-y$ view of the Outer Tracker.	33
2.14	An older version of the Serenity board.	34
3.1	The track-finding algorithm concept.	36
3.2	The L1 tracking efficiency as a function of η and the track z_0 resolution.	38
3.3	The seeding process using Virtual Modules.	38
3.4	A simplified flowchart of the track-finding algorithm.	43
3.5	An overview of the tracklet algorithm using only modules for the tracker barrel.	44

3.6	A flowchart of the VMR design.	52
3.7	An overview of the reduced configuration of the tracklet algorithm, the skinny chain.	54
3.8	The floorplan of the skinny chain on a Xilinx Virtex Ultrascale+ VU7P.	55
3.9	A simplified flowchart of the Combined Module version of the track- finding algorithm.	56
4.1	The p_T^{miss} trigger rates for different values of minimum track p_T . .	63
4.2	The jet p_T (left) and H_T (right) trigger rates for different values of minimum track p_T	63
4.3	The p_T^{miss} trigger efficiency.	64
4.4	The p_T^{miss} residual.	65
4.5	The jet p_T trigger efficiency.	65
4.6	The jet p_T residual.	66
4.7	The H_T trigger efficiency.	66
4.8	The H_T residual.	67
5.1	A Firefly transmitter and receiver pair connected to a Serenity board.	70
5.2	The general loopback setup and signal path for the Firefly optical tests.	72
5.3	The setup of a Serenity board with Firefly transceivers connected in a loopback.	72
5.4	An example bathtub curve.	73
5.5	Transmitter parameter scan results for the 4+4-channel Firefly. . .	78
5.6	Bathtub scans of all links on the 4+4-channel Firefly.	78
5.7	Parameter scan results of the transmitter settings for the 12-channel Beta parts.	79
5.8	Bathtub scans for the 12-channel Beta parts.	80
5.9	Receiver parameter scan results for the 4+4-channel Firefly.	81
5.10	Transceiver parameter scan results for ICjul22d with the CDR turned off.	82
5.11	Bathtub scan for ICjul22d with the CDR turned off.	83
5.12	The setup and signal path for the optical oscilloscope measure- ments of the Firefly transmitters.	83
5.13	An example eye diagram with the definitions of various optical signal characteristics.	84
5.14	Eye diagrams of all links of ICjul22c.	86
5.15	The transceiver OMA for all links of the 4+4-channel Firefly. . . .	86
5.16	The transceiver OMA for each 12-channel Beta part.	87
5.17	Extinction ratio for the 12-channel Beta parts.	88
5.18	Eye diagrams from the optical oscilloscope of channel 2 on ICJul22e.	89
5.19	Eye diagrams from the optical oscilloscope of two channels that no longer work.	90

5.20	The setup and signal path for the attenuation measurements. . . .	91
5.21	The setup of a Serenity board with Firefly transceivers connected in a loopback via an attenuator.	91
5.22	The BER as a function of applied attenuation on the 4+4-channel Firefly.	92
5.23	The BER as a function of applied attenuation on the 12-channel Firefly transceivers.	93
5.24	A comparison of how cross-talk in the receiver affects the attenu- ation measurements.	94
5.25	The extinction ratio and Tx OMA for the 12-channel Beta part ICjul22j.	95
5.26	A comparison of how cross-talk in the receiver affects the attenu- ation measurements with fixed attenuators.	96
5.27	The BER as a function of the calculated receiver OMA for the 4+4-channel Firefly.	97
5.28	The BER as a function of the calculated receiver OMA for each 12-channel Beta part.	98
5.29	The RSSI as a function of the calculated average power at the receiver for each 12-channel Beta part.	100
6.1	A Feynman diagram of the $\tilde{q}\tilde{q}$ signal model.	106
6.2	The H_T distribution for various background and signal events. . .	110
6.3	The double- b -tag region definitions for the btagHbb and ParticleNet taggers.	112
6.4	Distribution of the double- b -tag discriminator values for the two selected AK8 jets for a signal sample with $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV.	113
6.5	Distribution of the double- b -tag discriminator values for the two selected AK8 jets for QCD background events.	114
6.6	Distribution of the double- b -tag discriminator values for the two selected AK8 jets for $t\bar{t}$ background events.	114
6.7	Distribution of the double- b -tag discriminator values for the two selected AK8 jets for W +jets.	115
6.8	Distribution of the double- b -tag discriminator values for the two selected AK8 jets for Z +jets.	115
6.9	A comparison of the reconstructed AK8 jet masses.	116
6.10	Example of mass regions in the 2D jet mass plane developed for the btagHbb double- b -tagger and the Softdrop mass.	117
6.11	Distribution of the reconstructed jet mass for the two selected AK8 jets for various m_{H_1} masses.	119
6.12	Distribution of the reconstructed jet mass for the two selected AK8 jets for various m_{SUSY} masses.	120

6.13	Distribution of the reconstructed jet mass for the two selected AK8 jets for the QCD background that passed the kinematic event selection.	121
6.14	Distribution of the reconstructed jet mass for the two selected AK8 jets for the QCD background that passed the kinematic event selection and the double- b -tag selection.	122
6.15	Distribution of the reconstructed jet mass for the two selected AK8 jets for the $t\bar{t}$ background that passed the kinematic event selection.	123
6.16	Distribution of the reconstructed jet mass for the two selected AK8 jets for the $t\bar{t}$ background that passed the kinematic event selection and the double- b -tag selection.	123
6.17	Distribution of the reconstructed jet mass for the two selected AK8 jets for the W +jets background from events that passed the kinematic event.	124
6.18	Distribution of the reconstructed jet mass for the two selected AK8 jets for the Z +jets background from events that passed the kinematic event.	124
6.19	Distribution of the reconstructed jet mass for the two selected AK8 jets for the W +jets background from events that passed the kinematic event selection and the double- b -tag selection.	125
6.20	Distribution of the reconstructed jet mass for the two selected AK8 jets for the Z +jets background from events that passed the kinematic event selection and the double- b -tag selection.	125
6.21	The expected yields in the TR for each signal mass region after the event selections.	126
6.22	A comparison of the predicted yields and the observed yields in the TR of the QCD MC background.	129
6.23	A comparison of the predicted yields and the observed yields in the VR of the QCD MC background.	130
6.24	A comparison of the predicted yields and the observed yields in Run 2 data from 2018 in the VR.	130
6.25	The effect of the JMS and JMR uncertainties.	133
6.26	The ROC curve for a range of double- b -tag discriminator sum cuts.	134
6.27	The number of mistagged jets for a range of double- b -tag discriminator sum cuts.	135
6.28	The ROC curve for a range of double- b -tag discriminator cuts after all event selections.	136
6.29	An example of mass region bins for ParticleNet with two signal regions.	138
A.1	Transmitter parameter scan results for ICapr22.	144
A.2	Transmitter parameter scan results for ICmay22a.	144
A.3	Transmitter parameter scan results for ICmay22b.	145
A.4	Transmitter parameter scan results for ICmay22c.	145

A.5	Transmitter parameter scan results for ICmay22d	146
B.1	Parameter scan results of the transmitter settings for the 12-channel Beta parts on Serenity HNZN.	148
B.2	Bathtub scans for the 12-channel Beta parts on Serenity HNZN. . .	149
B.3	Parameter scan results for ICjul22d with CDR turned off on Serenity HNZN.	150
B.4	Bathtub scan for ICjul22d with the CDR turned off on Serenity HNZN	151
7.1	Distribution of the reconstructed jet mass for the two selected AK8 jets for various m_{H_1} masses of events that passed the kinematic event selection	153
7.2	Distribution of the reconstructed jet mass for the two selected AK8 jets for various m_{SUSY} masses for events that passed the kinematic event selection.	154

List of Tables

3.1	Specifications of two models of Xilinx Virtex Ultrascale+.	40
3.2	Number of bits for each quantity in the InputStub and AllStub memories.	49
3.3	Number of bits for each quantity in the VMStub memories read by the ME.	49
3.4	Number of bits for each quantity in the VMStub memories read by the TE for inner layer seeding.	50
3.5	Number of bits for each quantity in the VMStub memories read by the TE for outer layer seeding.	50
3.6	The input and output memory types connected to the VMR_L2PHIB module.	51
3.7	The resource usage for the VMR_L2PHIB module.	51
3.8	Preliminary resource usage for the barrel-only chain.	55
3.9	Number of bits for each quantity in the AllStub Inner memory read by the Tracklet Processor.	57
3.10	The input and output memory types connected to the VMR_L2PHIB module for the Combined Module chain.	57
3.11	The resource usage for the VMR_L2PHIB module for the Combined Module chain.	58
5.1	The Beta Firefly transceivers that were delivered to Imperial College London in 2022.	71
6.1	Squark/gluon pair-production cross sections.	106
6.2	Branching fractions for $H_1 \rightarrow b\bar{b}$ for various Higgs masses.	107
A.1	The Alpha3 Firefly transceivers that were delivered to Imperial College London in 2022.	143

Chapter 1

Background

1.1 Introduction

Our understanding of the Universe is based on the behaviour of a handful of minuscule particles and the forces that govern their interactions. These elementary particles are the building blocks of everything we see around us. Among the many theories that attempt to describe the behaviour of these particles, the most successful one to date is the Standard Model (SM) of particle physics. Experimental tests of these theories involve the use of particle detectors, like the CMS detector, which played a pivotal role in the observation of the Higgs boson, the final elementary particle to be discovered in the SM [1].

Despite the success of the SM, it is not a complete theory, and there are many other theories that seek to better explain the Universe, e.g. supersymmetric extensions of the SM. Supersymmetry (SUSY) predicts the existence of additional particles beyond those already discovered. The Next-to-Minimal Supersymmetric Standard Model (NMSSM) is one such theory. This chapter provides an overview of particle physics: both the SM and SUSY. Further discussions on the NMSSM can be found in Chapter 6, which explores an NMSSM scenario where additional Higgs bosons exist and one of them is similar to the Higgs already found, albeit lighter.

1.2 The Standard Model

The Standard Model (SM) is the most successful theory so far at describing the particles in our universe and their interactions. The SM explains three of the four fundamental forces: the electromagnetic force, the weak nuclear force, and the strong nuclear force. The last force, gravity, is not included. As shown in Figure 1.1, the elementary particles in the SM are divided into two groups: the fermions of which matter consists, and the bosons which include force carriers. These particles interact and build up our visible universe.

Heavy particles can decay into lighter ones if allowed by conservation laws such as conservation of momentum or electric charge. Thus any given particle may have several possible ways of decaying. Which decay mode a particle will take is unknown beforehand, but the probability of any decay mode, called its branching fraction, can be measured and predicted.

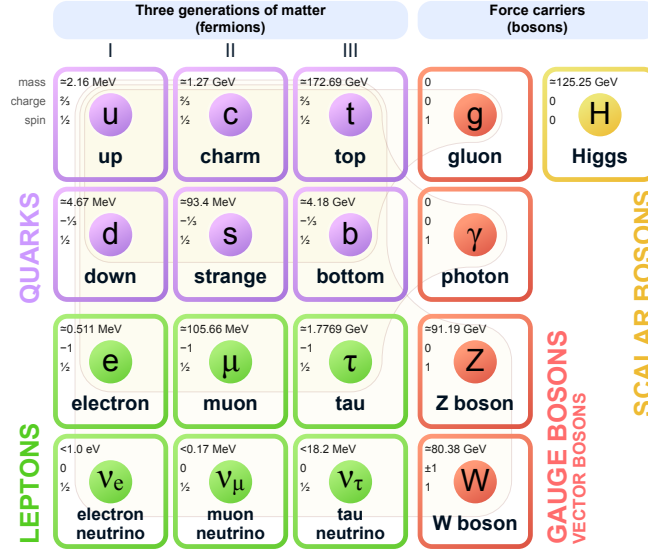


Figure 1.1: All elementary particles in the Standard Model [2]. The masses are from measurements and may change with more available data. The latest values can be found in Reference [3].

Mathematically, the SM is a quantum theory of fields. A field describes a physical quantity that for each point in time and space takes on some value. The field can be represented by a scalar or a vector. A simple example of a physical scalar field is temperature, which has a value at all points at all times. Particles in the SM can be described as excited states of their associated field. Fields may couple to other fields, which allows their corresponding particles to interact. More details of the SM can be found in References [4, 5], on which the following is based.

1.2.1 Fermions

Fermions have half-integer spin: a quantum number of intrinsic angular momentum which can only take on discrete values. According to the Pauli exclusion principle [6], two identical fermions cannot both be in the same state.

Within the SM, the fermions include particles called quarks and leptons. There are six types of quarks (u , d , c , s , t , b), where the types are referred to as flavours. There are also six types of leptons grouped into three flavours (e , ν_e , μ , ν_μ , τ , ν_τ). Each fermion has an anti-particle with the same mass but the opposite charge. The fermions are divided into three generations with

the lightest flavours in the first generation. If no conservation law forbids it and a suitable force couples them, the heavier particles tend to decay to the lighter ones. As a consequence, stable matter, such as atoms, generally consists of the particles in the first generation: up quarks (u), down quarks (d), electrons (e^-).

Quarks

Each generation of quarks contains an up-type quark with a $2/3$ positive electric elementary charge and a down-type quark with a $1/3$ negative elementary charge. Similar to particles having an electric charge due to the electromagnetic force, quarks also carry a colour charge, which is the corresponding charge of the strong force. Unlike the other forces, which become weaker at long distances, the strong force becomes stronger with the distance between coloured particles. Although quarks have either the colour charge red, green, or blue, only colourless particles can exist freely because particles prefer to be in the lowest energy state possible. For that reason, quarks must form bound states via strong interactions with other suitable quarks which results in composite particles called hadrons. This phenomenon is called colour confinement. There are two types of hadrons: baryons and mesons. Some of the most common baryons, such as the proton and the neutron, are created by combining three quarks with one of each colour. More rare baryons can comprise more than three quarks. Mesons are formed by taking a quark and an antiquark with one colour and its anti-colour.

Because of colour confinement, if a quark were to be created or forced out of e.g. a proton in a collision, the energy of the quark from the strong force would exceed the energy at which new particles can be created. The quark, therefore, causes a shower of other particles to ultimately form free colourless particles. This particle shower is called a hadron jet. Studying the jets can determine the properties of the original quark since quarks cannot be observed directly. For example, the momentum of the jet constituents should add up to the momentum of the original quark.

Leptons

Leptons can be electrically charged, such as the electron, or neutral such as the electron neutrino. There is one charged and one neutral lepton in each generation. Leptons are not subject to the strong force, as they do not carry any colour charge. Together with other particles, such as the proton and the neutron, charged leptons can create other composite particles such as atoms.

Neutrinos in the SM were originally assumed to be massless, however, observations of neutrino oscillations proved the opposite. Extensions of the SM are necessary to explain massive neutrinos, but their masses are yet to be determined.

1.2.2 Bosons

Interactions are described in the SM by particles exchanging bosons. All bosons have integer spin. Unlike the fermions, an unlimited number of indistinguishable bosons can reside in the same quantum state.

Gauge bosons

Gauge bosons mediate the fundamental forces. Photons (γ) are the mediators of the electromagnetic force which allows interactions between electrically charged particles. It is because of the electromagnetic force that light, electric fields, and magnetic fields exist.

The weak nuclear force, which acts on fermions, is responsible for flavour changes among the quarks and leptons and is mediated by the $W^\pm$ and Z bosons. These flavour changes are what causes e.g. radioactive decay. The weak force carriers interact with all particles that carry a weak charge, including themselves. The weak force can, along with the electromagnetic force, be described by the electroweak theory. The corresponding fields of this theory are the three weak isospin fields and the weak hypercharge field. The electromagnetic and weak charges are then represented as a mixture of the quantities of weak hypercharge and weak isospin.

The strong force is responsible for interactions between colour-charged particles and is mediated by gluons. The gluons bind quarks together to form colourless particles such as hadrons. Since gluons themselves carry a colour charge, they can self-interact. The strong interaction is described by the theory of Quantum Chromodynamics (QCD). All SM gauge bosons have spin-1.

Higgs boson

The last particle of the SM (discovered in 2012), is the Higgs boson [1]. It is a scalar spin-0 particle, the only SM particle with zero spin. The Higgs boson has no electric or colour charge and is not the cause of any fundamental force. Instead, the Higgs boson is a manifestation of the Higgs field which is responsible for giving other particles their mass through the Brout-Englert-Higgs mechanism. The Higgs mechanism is due to the Higgs field's non-zero value when it is in its lowest energy state is not zero, or vacuum expectation value.

The simplest model of the Higgs field ϕ has four degrees of freedom and consists of two complex scalar fields in a weak isospin doublet, namely

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (1.1)$$

In the SM, particles are excited states of quantum fields, thus electroweak gauge bosons arise from the electroweak fields, comprising of the weak hy-

percharge boson (B) along with three weak isospin bosons (W_1, W_2, W_3). However, these bosons are not observable because of the Higgs mechanism. The Higgs field's non-zero vacuum expectation value causes the symmetry of the electroweak field to spontaneously break, resulting in the electroweak gauge bosons mixing with three of the four degrees of freedom of the Higgs field to produce the massive $W^\pm$ and Z bosons. The remaining degree of freedom becomes the Higgs boson and its mass is related to the vacuum expectation value of the Higgs field. The γ is a mixture of the electrically neutral B and W_3 and does not gain any mass. Fermions gain their masses through so-called Yukawa interactions with the Higgs field.

1.2.3 Physics Beyond the Standard Model

Even though the SM has successfully described and predicted many experimental observations, it is established that it is not the final theory of everything. One clear example is that the model does not include gravity. Additionally, cosmological measurements indicate that there is more mass in the Universe than what can be directly seen, implying the existence of an unknown type of matter which has not yet been discovered, so-called dark matter. The SM does not contain any sensible dark matter candidates.

The SM does not describe why the Higgs boson mass is around the electroweak scale of 100 GeV. The Higgs boson mass is a free parameter, but the Higgs mass measured by experiments also receives contributions from quantum loop interaction corrections, where the Higgs creates and absorbs virtual particles. Examples of such loop corrections for the Higgs boson can be seen in Figure 1.2. These corrections increase quadratically with the mass scale at which the theory is valid. Therefore, if the SM is to remain valid at high mass scales, e.g. the Grand Unified Theory (GUT) scale of the order of 10^{16} GeV where the electroweak force and strong force could potentially be merged into a single force, the Higgs mass requires very precise adjusting of free parameters, or fine-tuning, to achieve the observed mass value. This is called the hierarchy problem. A theory is deemed more natural when less fine-tuning is required.

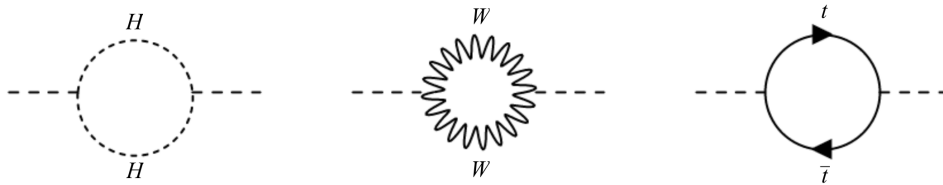


Figure 1.2: Examples of Feynman diagrams of quantum loop corrections for the Higgs boson mass. A dashed line represents a Higgs boson; a wavy line represents a gauge boson, in this case, a W -pair; and a solid line represents a fermion, in this case, a $t\bar{t}$ -pair.

A plethora of new models have been proposed in an attempt to explain the mysteries of the Universe, but none have yet been confirmed experimentally. One example is Supersymmetry (SUSY) [7], described further in the next section, Section 1.3. The search for evidence of physics Beyond the Standard Model (BSM) that might lead us to the desired theory of everything is an ongoing process. Precise measurements of rare decays or interactions are one approach to detecting BSM physics. If the decays are well-predicted in the SM, then the decay rate might indicate new physics if the measured rate differs from the predicted [8]. Another approach is to continue the measurements of the Higgs boson and its interactions with matter. The Higgs boson is fundamental for understanding the origin of mass, so even if unknown massive particles do not interact via any of the three fundamental forces, they could still interact with the Higgs boson. Such a BSM particle would affect the Higgs mass, via processes analogous to the quantum loop corrections shown in Figure 1.2, thus BSM particles must be compatible with the already observed Higgs mass. Any deviation from the SM could also be revealed by measuring the Higgs coupling to other particles which are sensitive to new physics [9], e.g. the top quark, making the Higgs sector a potential gateway to uncovering new BSM physics [8].

1.3 Supersymmetry

Supersymmetric extensions of the Standard Model are popular theories where a spacetime symmetry between fermions and bosons is incorporated [7]. As a result, supersymmetric partners (super-partners) to each particle in the SM are added to the theory. All particles have the same quantum numbers, for example, colour charge and electric charge, as their super-partners except for the spin. Hence SM fermions have supersymmetric bosons as their partners, and vice versa.

In some simple theories with unbroken symmetry, the super-partner would have the same mass as the original particle and the quantum loop corrections of the super-partners (similar to Figure 1.2) would exactly cancel out the quadratically divergent SM contributions to the Higgs mass, thus solving the hierarchy problem. These theories were considered natural, but are clearly insufficient, as many of the SUSY particles should have been discovered by now since the cross section of such particles would be large enough for the particles to be created at current experiments. It is possible to add a term in the definition of the theory which breaks this symmetry of equal masses between the particles and their super-partner. Therefore, most SUSY theories are supersymmetry-breaking, which leads to heavier super-partners compared to their SM counterparts. In the broken theory, the Higgs mass contributions are dependent on the squared mass of the heaviest SUSY particle. However,

SUSY models are still considered natural provided that the SUSY masses are small, i.e. around 1 TeV or less, as this would require less fine-tuning.

Many of the proposed theories suggest that the Lightest SUSY Particle (LSP) is stable, weakly interacting, and electrically neutral, which makes it a possible candidate for dark matter. To ensure that the LSP cannot decay to any SM particle, which is assumed in many SUSY models to avoid proton decays, a quantity called R -parity is introduced. R -parity is defined as $(-1)^{3B+L+2s}$ and is a symmetry related to the spin s , as well as the baryon number B and lepton number L . The baryon and lepton numbers are quantum numbers defined as the number of baryons (the number of quarks divided by three) and leptons, respectively. All SM particles have an R -parity of 1, whereas SUSY particles have an R -parity of -1. If the R -parity is conserved, the LSP is stable.

The Minimal Supersymmetric Standard Model (MSSM) is a SUSY model that introduces the fewest number of additional particles and states required to create a consistent theory compatible with known physics. If only the corresponding super-partners to the SM particles were added, the super-partners of the Higgs, called higgsinos, would create an anomaly caused by three gauge bosons coupled to an internal fermion, as illustrated in Figure 1.3. These triangular interactions cause calculations of the theory to diverge, which is unphysical. The fermions within the SM cancel out the possibility for such interactions exactly, but the additional higgsinos from a single Higgs field do not. Adding a second Higgs doublet, i.e. more higgsinos, could cancel the anomaly completely. Therefore, to avoid any theory inconsistencies, the Higgs sector in the MSSM consists of two complex Higgs doublets. One doublet gives mass to the up-type quarks, and the other gives mass to the down-type quarks and leptons. This results in a total of five physical Higgs bosons, including two SM-like Higgs bosons, where one of them can be chosen to be the Higgs boson already observed. There are also charged Higgs bosons. Despite numerous searches for MSSM particles, there is no experimental evidence of their existence. Lower bounds have been set for the masses of the theorised SUSY particles, such as the super-partners of quarks and gluons, which now have a limit greater than 1 TeV [3]. As a result, a large portion of the parameter space for the MSSM which was considered natural has been ruled out [10]. Many theories, such as the Next-to-Minimal Supersymmetric Standard Model (NMSSM), are extensions of the MSSM with different features which, in some cases, can solve the problem with unnaturality. The NMSSM is discussed further in Section 1.3.2.

1.3.1 Supersymmetric Particles

Below is a summary of various types of SUSY particles commonly found in SUSY models. The number of particles and which particles are predicted to exist depends on the theory. The ones listed here are found in all SUSY

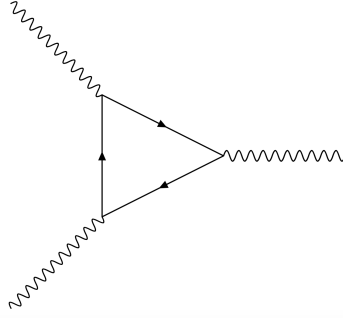


Figure 1.3: A triangle loop Feynman diagram of three gauge bosons connected to a fermion. These interactions cause theory anomalies.

theories, as they correspond to the super-partners of the SM particles, which is the bare minimum for a consistent theory. An additional particle appearing in the NMSSM, the singlino, is also described. While some of the particles listed here may not appear as physical entities, their mixing with other particles leads to physical mass eigenstates which could be detected.

Sfermions

Sfermions are the spin-0 bosonic super-partners of the SM quarks and leptons, referred to as squarks ($\tilde{q}$) and sleptons ($\tilde{l}$). Using the naming convention for the supersymmetric partner of fermions where the letter *s* is added at the beginning of the name, the super-partner of the up quark is the sup squark ($\tilde{u}$), and the partner of the electron is the selectron ($\tilde{e}$), and so forth.

Gauginos

Gauginos are the super-partners of gauge fields. All gauginos have a half-integer spin, most often $1/2$, thus they are fermionic. The naming convention for super-partners of gauge fields is to add a suffix *-ino* to the name.

Gauginos include particles such as gluinos ($\tilde{g}$) which are the super-partners of the gluon, along with electroweak gauginos called binos and winos. A bino is the super-partner of the weak hypercharge B boson, while winos are the super-partners of the weak isospin fields.

Higgsino

Higgsinos ($\tilde{H}$) are the spin- $1/2$ super-partners of the Higgs bosons. There are both electrically neutral and charged higgsinos, as SUSY extensions of the SM have additional electrically charged Higgs bosons and the electric charge of the super-partner matches its super-partner. The higgsino is not a physical particle in most SUSY theories, thus cannot be observed. Instead,

the higgsinos mix with other electroweak gauginos to create physical particles as a result of electroweak symmetry breaking.

Chargino and Neutralino

Thanks to the nature of particle mixing, the binos, winos, and higgsinos can in general mix to create mass eigenstates. These physical particles are called charginos and neutralinos depending on whether they are electrically charged or neutral, respectively. The charginos are denoted $\tilde{\chi}_i^\pm$ and the neutralinos are denoted $\tilde{\chi}_i^0$, where $i \in \mathbb{N}$ and $i = 1$ refers to the lightest such particle.

Singlino

In some SUSY extensions, like in the NMSSM described further in Section 1.3.2, a weak isospin singlet scalar field is added to the model which results in additional Higgs bosons and a so-called singlino. The singlino, a type of neutralino, is the fermionic spin-1/2 component of the additional superfield [10]. The singlino only couples to other neutralinos and Higgs bosons, thus making them very hard to detect and a potential candidate for dark matter.

1.3.2 Next-to-Minimal Supersymmetric Standard Model

The MSSM offered a solution to the problematic divergent quantum loop corrections of the SM, and it was considered natural for low SUSY masses. However, experimental findings have significantly constrained the regions of the MSSM parameter space where the model was deemed natural. It is worth noting that the SUSY masses could potentially be beyond the reach of current experiments, thus the model has not been ruled out. Nonetheless, there are additional challenges and concerns associated with the MSSM. For example, a mass term with a parameter μ appears in the mathematical formulation of the model: $\mu H_u H_d$. The μ parameter, which is a free parameter, couples the two Higgs doublets H_u and H_d . This term is involved in giving the Higgs bosons (and higgsinos) their masses, thus phenomenology requires μ to be of the order of the electroweak scale. However, there is no theoretical reason as to why μ should be close to the electroweak scale, and now the theory depends on both the electroweak scale and the SUSY breaking scale. The NMSSM solves the dependency on two energy scales by extending the MSSM with the introduction of a singlet scalar superfield (S) that replaces μ in the term which couples the two scalar Higgs doublets: $\lambda S H_u H_d$, where λ is a coupling constant [11]. The mass term is then naturally generated at an electroweak scale by this coupling, as the singlet field's vacuum expectation value is of the right scale. This leads to the theory only being dependent on the SUSY breaking scale.

Because of the extra superfield, the NMSSM has an additional three particles compared to the MSSM: the singlino and two Higgses. As mentioned, a

singlino-like particle is also a promising dark matter candidate. In the scenario where the singlino is the stable LSP, it would end up as missing energy inside particle detectors, as the singlino is neutral and weakly interacting [10]. Singlinos and additional Higgses, among other NMSSM particles, could theoretically be created in particle colliders such as the LHC, but no significant evidence for such particles has yet been observed [12]. A search for light Higgs bosons from supersymmetric decay cascades where the singlino is the LSP is presented in Chapter 6.

Chapter 2

The LHC and the CMS Detector

2.1 Introduction

To study the SM and its particles, big accelerators like the Large Hadron Collider (LHC) are built to collide e.g. protons at high energies. The collisions create various particles that are measured using gigantic detectors. One of the main detectors at the LHC is the Compact Muon Solenoid (CMS) detector, which has been operational for over a decade and contributed to the detection of new particles and interactions, including the Higgs boson. However, upgrades are necessary to be able to measure more rare events at a reasonable timescale, which is why the upcoming High-Luminosity LHC (HL-LHC) project will increase the number of simultaneous particle collisions up to 200 [13]. This chapter describes the LHC and its HL-LHC upgrade, as well as the CMS detector.

Because of the sheer amount of data generated by hadron collider experiments, only the most interesting collisions are saved. To select these events, so-called triggers are used which rapidly measure quantities, such as the total energy involved in the collision, and determine whether to keep the event based on a predefined threshold. CMS uses two triggers, the first of which is the hardware-based Level-1 Trigger (L1T). The L1T will need an upgrade for the HL-LHC to handle the increased number of collisions. One improvement is to use the trajectories of all charged high-momentum particles as input for the trigger selection. This has never been done in the CMS detector before. The particle trajectories, or tracks, enable better resolution of where a collision originated, which is critical to separate the simultaneous collisions. Tracks will also make it possible to use more sophisticated trigger algorithms in the L1T. The L1 track-finding algorithm, i.e. the algorithm which will find and reconstruct the tracks from the data, is explored in Chapters 3-4.

While the detector measures the particle collisions, the back-end systems, which include the L1T, control the data acquisition and communication with the detector. The back-end systems are located away from the detector, and a diagram of the data-flow can be seen in Figure 2.1. As part of the CMS upgrade, improvements are being made to the back-end hardware which consists of a multitude of processing boards. Data is transmitted between boards via optical fibres, and optical transceivers are responsible for converting the boards' electrical signals into optical signals and vice versa. Given the increase of data transmitted in the CMS upgrade, it is crucial to employ high-speed optical transceivers capable of handling the demands. Performance measurements of new 25 Gb/s optical transceivers envisaged for the CMS upgrade are discussed in Chapter 5.

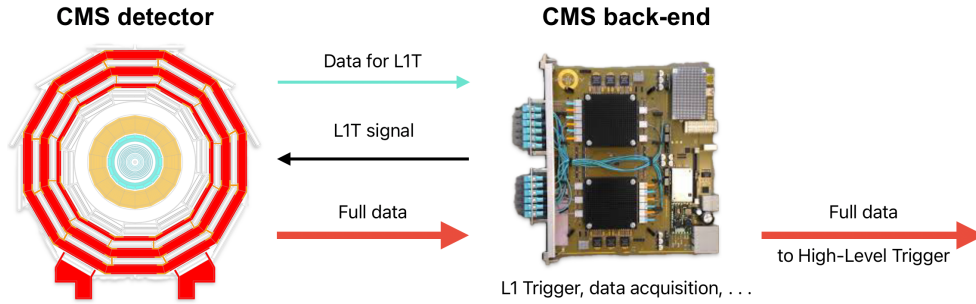


Figure 2.1: A simplified illustration of the data-flow between the CMS detector and the back-end systems. The full data does not get transmitted unless a positive signal from the L1T has been received by the detector.

2.2 The Large Hadron Collider

With its 27 km circumference, the LHC [14] is the largest particle accelerator in the world and has contributed to many important results in particle physics throughout its active years since 2008.

The circular structure of the LHC contains two parallel beamlines that accelerate particles in opposite directions. It can currently accelerate protons up to an energy of 6.8 TeV [15], corresponding to a velocity very close to the speed of light ($0.999\,999\,99\,c$). At the interaction points located at the centre of the four main detectors, see Figure 2.2, the paths of the two beams are brought together. When a collision occurs, the gluons and quarks inside the protons interact to create particles different from the original ones. The higher the energy involved in the collisions, the heavier particles can be produced. The LHC's 13.6 TeV collisions can, therefore, create rare or possibly even currently undiscovered particles. The probability that two particles will collide and a certain process will take place is given by the cross section (σ).

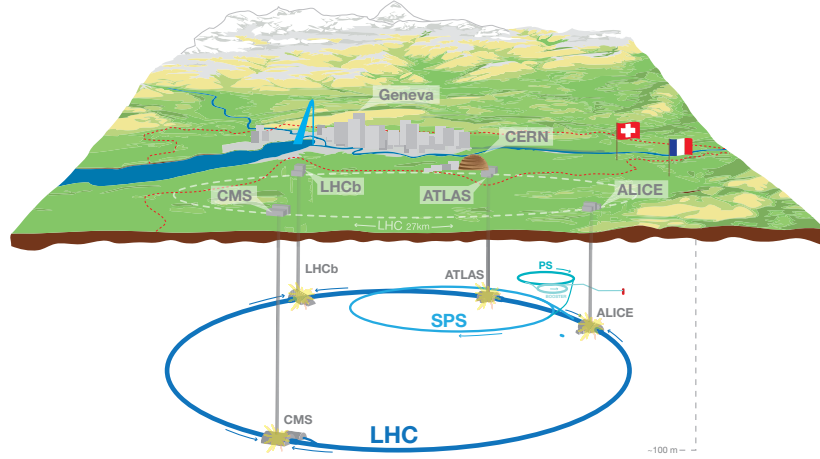


Figure 2.2: The LHC and the locations of the four main detectors: ALICE, ATLAS, CMS, and LHCb [16]. The smaller circular structures are the Super Proton Synchrotron (SPS) and the Proton Synchrotron (PS), which are some of the pre-accelerators for the LHC.

The performance of a particle accelerator can be described by the instantaneous luminosity which is proportional to the number of collisions at a given time. Integrated luminosity gives a measure of the total number of collisions recorded over a time period. Collisions at the LHC are usually performed using protons. The protons are collected into bunch trains, with a minimum spacing of 25 ns between the bunches [8]. Each bunch contains approximately 10^{11} protons, but the number of collisions per bunch-crossing is a function of luminosity which varies depending on the settings of the LHC. Processes where particles collide and undergo a large momentum transfer are sometimes referred to as hard scattering. Hard scatterings are highly interesting collisions where rare or even new particles can be created, however, most interactions are so-called minimum bias events. The simultaneous proton-proton collisions within a bunch-crossing, or event, that is not a hard-scattering collision are referred to as pileup.

The data collection at the LHC is split up into runs which last a few years each. The integrated luminosity and peak instantaneous luminosity that were delivered to the CMS experiment can be seen in Figure 2.3. At the start of Run 1 (2010-2012), the average number of interactions per bunch-

crossing ($\langle\mu\rangle$) produced by the LHC was 10, with an instantaneous luminosity somewhat lower than the nominal $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. This increased to a peak instantaneous luminosity of $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ for Run 2 (2015-2018), yielding an average of $\langle\mu\rangle = 34$. In Run 3 that started in 2022, the LHC provided a $\langle\mu\rangle \approx 44$, again with a peak luminosity of around $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

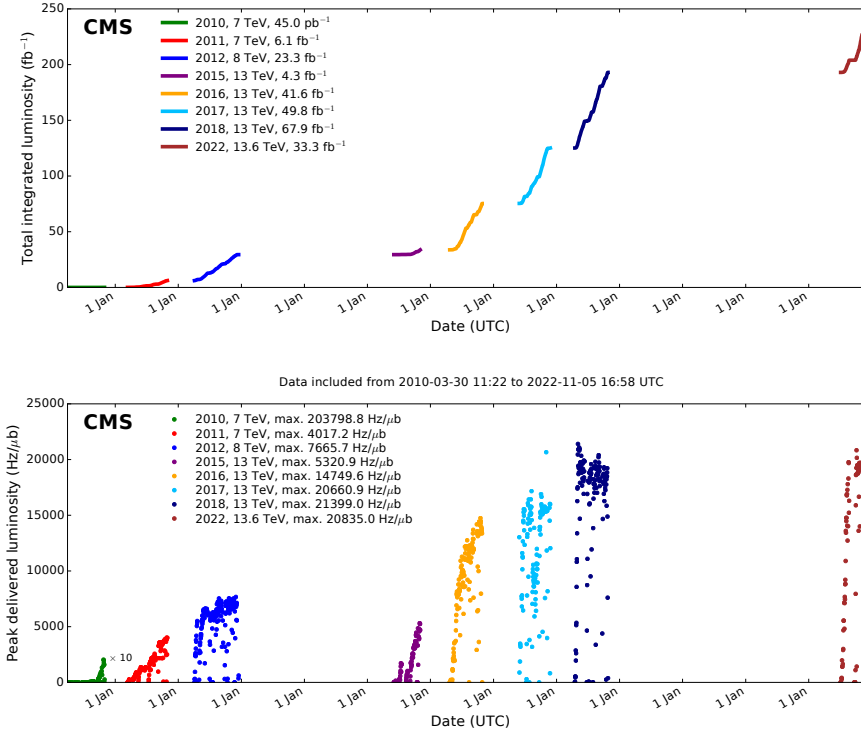


Figure 2.3: The integrated luminosity (upper) and instantaneous luminosity (lower) delivered at the CMS experiment for each year [17]. Note that the instantaneous luminosity for 2010 has been magnified by a factor of 10.

Since the LHC and its detectors started operating fully in 2010, the LHC has delivered an integrated luminosity of approximately 190 fb^{-1} . Around 30 fb^{-1} were delivered at 7 TeV and 8 TeV during Run 1, and around 160 fb^{-1} were delivered at 13 TeV during Run 2 [18]. Out of the data recorded by CMS during Run 2, 139 fb^{-1} are considered good for physics analysis. Currently, Run 3 is ongoing with a beam energy of 13.6 TeV. This run will last until 2025 [19] with the aim of collecting an additional 250 fb^{-1} of data [20].

A high instantaneous luminosity implies more interactions, hence more hard scattering events, but can also lead to a greater number of pileups which could interfere with the measurements of the hard scattering collision. This is something the detectors at the LHC need to cope with. The spatial spread of a bunch-crossing is normally a few centimetres in the direction of the beam-

line [14], and because all collisions do not happen at the same point, it is possible to separate which particle came from which interaction with high-resolution detectors.

2.3 The CMS Detector

The CMS detector is one of the main detectors at the LHC. Together with ATLAS, it is one of the two general-purpose LHC detectors that can search for a broad range of physics and new phenomena. The main goals of the detector include comprehensive studies of the Higgs boson and its properties as well as searches for evidence of BSM physics, such as SUSY. The detector is located about 100 m underground in the so-called detector cavern where all the particle collisions occur. A separate service cavern that is shielded from radiation is located next to the detector cavern, containing items such as readout electronics and gas supplies needed by CMS. The detector back-end electronics in the service cavern communicate with the detector and perform the initial processing of the data. The final processing of the data takes place above ground to identify the particles created in the collisions and determine e.g. their momenta and trajectories. A brief overview of the detector and its subsystems is given below, but further details can be found in the CMS technical report, Reference [21].

Built around a huge solenoid magnet, the CMS detector consists of several subsystems surrounding the interaction point in various barrel-like layers with endcaps on each side of the barrels, see Figure 2.4. CMS is designed to determine the tracks and measure the energies of particles from the high-energy collisions created by the LHC. The powerful solenoid significantly bends charged particles, even high-momentum muons, for precise momentum measurements. How different particles travel and interact in the various subdetectors can be seen in Figure 2.5. The detector barrels detect particles travelling at large angles with respect to the beamline, and the endcaps detect particles travelling at a small angle from the beam (also called the forward direction). To describe the direction and position inside the detector, the beamline is defined as the z -axis, with the origin located at the centre of the detector. The x -axis is pointing towards the centre of the LHC and the y -axis is pointing upwards. The polar angle from the z -axis is denoted θ , but is frequently represented by the pseudorapidity η defined as

$$\eta = -\ln \left(\frac{\theta}{2} \right). \quad (2.1)$$

Hence, large absolute values of η correspond to the forward region, and $\eta = 0$ corresponds to the direction perpendicular to the beam. The radial distance from the z -axis is denoted r , and the azimuthal angle perpendicular to the z -axis is denoted ϕ . The transverse plane is defined as the $r - \phi$ ($x - y$) plane,

i.e. perpendicular to the beam direction, and the momentum component in this plane is denoted p_T . Some particle types can not be measured directly in the detector, such as neutrinos. Instead, their presence and trajectories are constrained by the missing transverse momentum defined as

$$p_T^{\text{miss}} = - \left| \sum_i \vec{p}_T(i) \right|, \quad (2.2)$$

where $p_T(i)$ is the transverse momentum for particle i in the event.

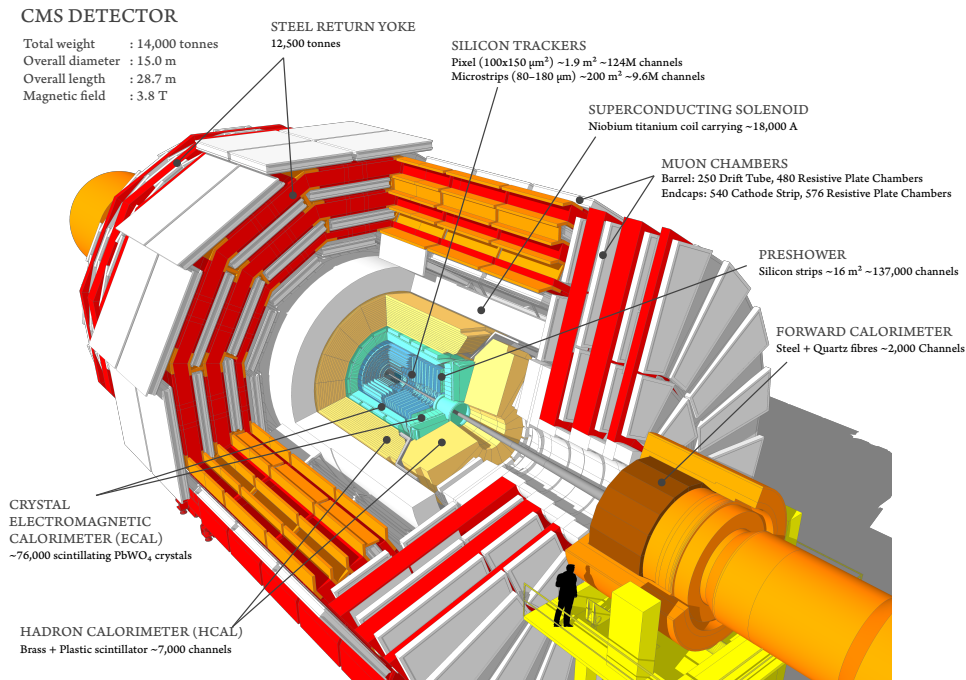


Figure 2.4: The CMS detector and the location of its subsystems [22]. Collisions occur at the centre of the silicon trackers.

2.3.1 Solenoid Magnet

CMS's solenoidal magnet creates a magnetic field up to 4 T that is uniform inside the magnet volume. Charged particles' trajectories bend in a magnetic field, and because of the uniform magnetic field, their path will follow a helix. The track component parallel to the beamline is unaffected by the magnetic field, thus the charged particles will only bend in the plane that is perpendicular to the beamline. The projection of the track in the $r - \phi$ plane is a circle,

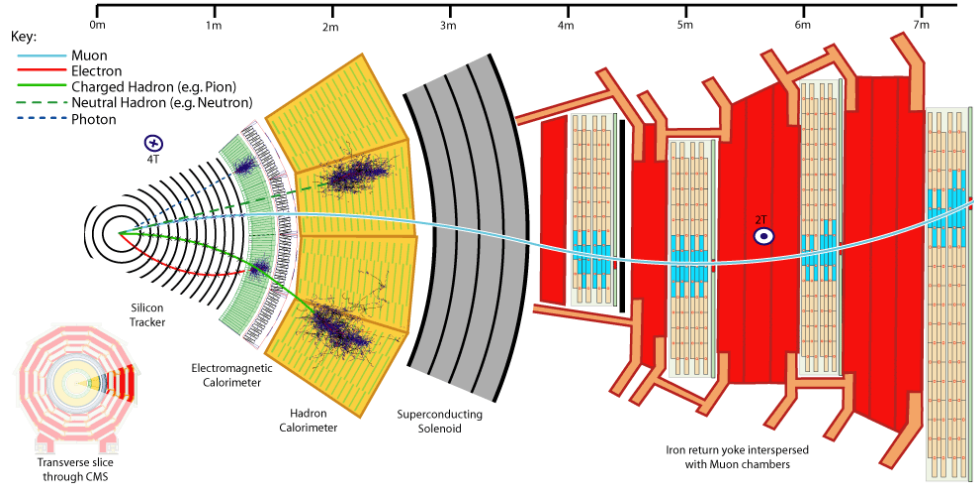


Figure 2.5: A slice of the CMS detector showing how different types of particles interact with the various subdetectors [23].

and the radius R (in centimetres) of the helix can be described as

$$R = \frac{p_T}{0.003qB}, \quad (2.3)$$

where p_T is given in GeV, q is the charge of the particle (in units of the elementary charge), and B is the magnetic flux density in T inside the CMS solenoid. The radius depends on both the p_T of the particle and the charge, hence the more momentum a particle has, the less it will bend. The sign of the particle's charge will dictate which direction it will bend. Because of scattering when particles interact with matter, subdetectors like the tracker and the calorimeters are placed inside the solenoid for better resolution.

Outside the solenoid magnet, there are return yokes made of steel to control the magnetic field and to stop all SM particles apart from muons and neutrinos from entering the muon chambers. The return yokes also act as the main support system of the whole experiment.

2.3.2 Tracker

The detector closest to the interaction point is called the tracker. It tracks charged particles in three dimensions and measures their charge as well as momentum thanks to the large magnetic field parallel to the beam axis. Since it is easier to measure the kinematics of particles with a large bend, the track resolution increases the lower the momentum. Good momentum resolution is important for e.g. p_T^{miss} measurements which could involve exotic particles undetectable for CMS. The tracks can subsequently be used to reconstruct the point at which the collisions occurred, also called the primary vertex,

which is also a critical component for later analysis of the interactions. This is possible because of its proximity to the interaction point and excellent sensor resolution. The tracker is built to interact as little as possible with the particles to minimise particle scattering which degrades the resolution. The tracker consists of the pixel and the strip silicon detectors, where the former is located closest to the beamline and has higher granularity than the latter. The tracker covers the region $|\eta| < 2.5$. A cross section of the layout of the tracker which will be used until the end of year 2025 (also known as Phase-1) can be seen in Figure 2.6. Because of the shape of the endcap layers, they are sometimes referred to as disks. The tracker was originally designed to operate with an average pileup of 20-30 collisions per bunch-crossing but is performing well even at higher luminosities [24].

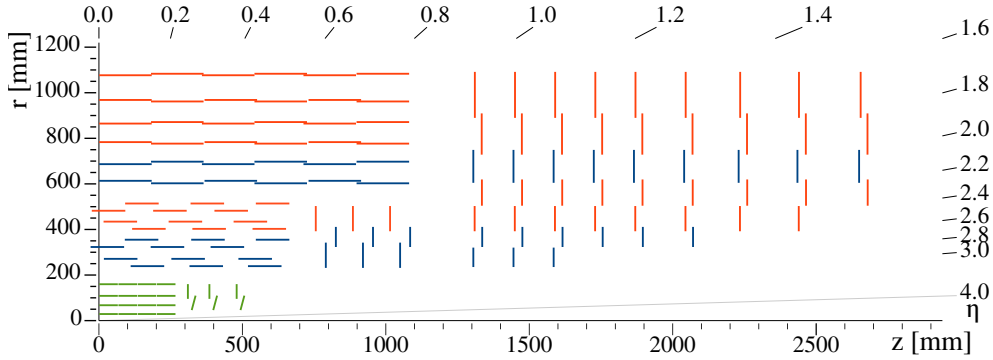


Figure 2.6: The Phase-1 tracker layout [24]. It shows a quarter of the tracker in the $r - z$ plane, with the interaction point at the origin. The pixel detector is depicted in green, and the single- and double-sided strip detector is shown in red and blue, respectively.

To not interfere with the particle trajectory the tracker is relatively light-weight, thus the particles can continue to the calorimeters with minimal multiple scattering. Charged particles will ionise the tracker's silicon sensors, which leads to a small charge that is measured. However, electrically neutral particles pass undetected through the tracker.

Pixel Detector

As previously mentioned, the granularity of the pixel tracker is the highest within all of CMS to cope with the small distances between particles when close to the interaction point. It can even measure tracks and secondary vertices coming from short-lived particles that in turn decay into other particles inside the tracker volume, such as hadrons containing b quarks.

Because of the harsh radiation environment closest to the interaction point, the pixel detector was completely exchanged and upgraded to the current Phase-1 pixel detector during the 2016/2017 year-end stop. Each layer consists

of radiation tolerant silicon sensors with the four barrel layers placed 30-160 mm away from the beamline [25], as seen in Figure 2.6. In total, there are 1856 sensor modules and over 100 million pixels. The layout is designed to yield four hits per charged particle in the region $|\eta| < 2.5$. At least three hits are needed to reconstruct a track (two if the particle is assumed to originate from the centre of the detector). The $100 \times 150 \mu\text{m}^2$ pixel size yields good resolution in both z and ϕ , and the average occupancy, or hits per pixel, is of the order of 10^{-4} per pixel per bunch-crossing at the nominal luminosity [21]. The overall spatial resolution of the pixel tracker is of the order of $10 \mu\text{m}$.

The silicon sensor modules also contain readout chips. These chips amplify the sensor signals and temporarily store the recorded events in buffers. Almost 200 events can be stored simultaneously. If a signal is received from the back-end systems to further process the event (Section 2.3.6), the data is transmitted via optical fibres from the detector front-end to the data acquisition system in the back-end.

Strip Detector

The strip detector is located at radii larger than 200 mm in the barrel, surrounding the pixel tracker, thus increasing the overall volume for particle tracking. The strip detector consists of 15 148 sensor modules spread over ten barrel layers and twelve disks on either side of the barrels. As it is located further away from the interaction point, the resolution and granularity are not required to be as high. Instead, the sensors consist of silicon strips, which have a coarser resolution in the z -direction (r -direction) in the barrel (disk). This allows the resolution in the $r - \phi$ plane to remain excellent. By glueing two single-sided strip modules together at an angle, it is possible to acquire a decent resolution in the $r - z$ plane. Thus, some of the strip detector layers consist of these double-sided modules. In total, there are about 10 million strips. The minimum strip size goes from $10 \text{ cm} \times 80 \mu\text{m}$ to a maximum strip size of $25 \text{ cm} \times 180 \mu\text{m}$. For the double-sided modules, the single-hit resolution can go down to $23 \mu\text{m}$ depending on the sensor module and its location. Even with the larger sensor cells, the occupancy is still at most a couple of per cent.

The strip sensor modules also contain similar readout chips as for the pixel tracker.

2.3.3 Electromagnetic Calorimeter

The energy of electrons and photons is measured in the Electromagnetic Calorimeter (ECAL) with high precision. The ECAL does so by completely stopping the electrons and photons using dense crystals made of lead tungstate. Lead tungstate crystals are transparent and when particles which can interact via the electromagnetic force travel through them, scintillation light is created inside the crystals. When an electron or a photon traverses the lead tungstate

crystals, they turn into electromagnetic showers: a high-energy photon loses energy in the dense material by creating electron-positron pairs, as long as it is above the energy required for pair production. An electron or positron in the material will in turn emit photons (bremsstrahlung), thus leading to a cascade of particles. The amount of light produced, which is related to the energy of the electron or photon, can then be measured using photodiodes which convert the light into an electrical signal for readout.

The ECAL consists of a single barrel layer with over 60 000 crystals ($2.18 \times 2.18 \times 23$ cm) and a single endcap on each side of the barrel with 15 000 crystals ($2.85 \times 2.85 \times 22$ cm), covering the region $|\eta| < 3.0$. Just before the endcaps is a preshower detector made out of lead and silicon sensors. Photons will create electron-positron pairs when travelling through the lead layer, which can then be measured by the silicon sensors. The preshower detector increases the spatial resolution of the ECAL, which otherwise has a limited spatial resolution, to separate interesting high-energy photons from close pairs of low-energy photons in a region that is not fully covered by the tracker.

Heavier particles, such as hadrons and muons, deposit a small amount of energy in the ECAL but are not stopped, thus continuing further out of the detector.

2.3.4 Hadron Calorimeter

As suggested by the name, the Hadron Calorimeter (HCAL) accurately measures the energy of hadrons. Similar to the ECAL, it does so by being built from dense absorber materials, such as brass, to stop the hadrons through hadronic showers. The hadronic showers are caused by hadrons colliding with the nuclei of the absorber material, which creates both additional hadrons and electromagnetic cascades. The showers are detected using photodetectors connected to plastic scintillator tiles placed in alternating layers between the absorber material. The HCAL covers the region $|\eta| < 3.0$.

Directly outside the solenoid in the barrel region, there is an additional Hadron Outer (HO) detector. The HO increases the effective thickness of the HCAL by measuring the energy from high-energy particles that might have leaked through the main calorimeter, thus improving the energy resolution. To assist the hadron endcap calorimeters, a forward calorimeter (HF) has also been installed outside the muon detectors in the forward region. Positioning the HF further away from the interaction point increases the coverage of the HCAL detector up to $|\eta| < 5$, which is important for p_T^{miss} measurements.

2.3.5 Muon Detectors

Muons are the only SM particles aside from neutrinos that are stopped neither by the ECAL nor the HCAL. Located outside the solenoid, the muon detectors aim to identify and accurately measure the momenta of muons. They consist

of four subsystems, all of which use gaseous detector technologies: the drift tubes (DTs), the cathode strip chambers (CSCs), the resistive plate chambers (RPCs), and gas electron multipliers (GEMs). The systems are located in different areas of the detector in various layers, and each subsystem uses different detector techniques which create precise and robust measurements of the muons. The coverage of the four muon subsystems combined is $|\eta| < 2.4$.

The muons are not stopped by the muon detectors, thus escaping the CMS detector.

2.3.6 Trigger System

Collisions at the LHC take place with a minimum interval of 25 ns, equating to a maximum potential output data rate of 40 MHz. CMS can save data to disk at a rate of 1 kHz (with an average event size of 2 MB) [26], thus it is not possible to keep all data. There is also a limitation on the rate at which data can be transmitted or received, i.e. bandwidth, which dictates how many events can be transmitted from the detector. Triggering systems decide which events are selected for final storage by calculating quantities which may imply interesting physics, such as the momentum of particles or the p_T^{miss} . If one of the various quantities reaches above a certain threshold, the processing of the event will continue. A large value of p_T^{miss} suggests that particles invisible to the detector were created. Many BSM models include undetectable particles, thus p_T^{miss} is an important quantity to measure. Another widely used quantity is the total jet transverse momentum H_T , defined as

$$H_T = \sum_j |\vec{p}_T(j)|, \quad (2.4)$$

where $p_T(j)$ is the transverse momentum for jet j in the event. A large value of H_T can imply high-energy jets, a characteristic of many physics studies. This quantity offers valuable insight into the total energy involved in the collision, which plays a crucial role in the identification of events that warrant further examination or storage for subsequent analysis.

The CMS trigger system consists of two stages for an efficient selection of events: the hardware-based Level-1 (L1) trigger [27] and the software-based High-Level Trigger (HLT) [28].

L1 Trigger

The Level-1 Trigger (L1T) is the first part of the CMS trigger system. The detector readout electronics have limitations on the rate at which events can be transmitted and the optical links used to transmit the detector data have insufficient bandwidth for a 40 MHz readout, thus the L1T is needed to decrease the rate to 100 kHz. The L1T uses high-speed electronics and reduced resolution data from the calorimeters and the muon detectors to make a trigger

decision. Latency is the time it takes for a system to respond to a given request or input. The detector readout electronics can temporarily store around 200 events in their buffers, thus the readout chips need to have received a positive decision, called a L1 accept (L1A), within a latency of $4\text{ }\mu\text{s}$ for the event to be further considered [29]. In other words, the latency must be less than the length of the buffer divided by the collision rate (40 MHz). If the trigger is fired, the L1A is sent to the front-end electronics, and the full data from all subdetectors is transmitted via the Data Acquisition (DAQ) system to the HLT for further processing.

Only around $1\text{ }\mu\text{s}$ of the total latency is dedicated to the L1T calculations [21]. The remaining time goes to transmitting the data and the L1A between the detector and the L1T located in the service cavern underground. Signal transmission is done using high-speed optical links. Because the L1T has to be very fast, its algorithms are run on custom processor boards equipped with Field Programmable Gate Arrays (FPGAs). FPGAs are very fast programmable chips, thus they can be updated and reconfigured when needed. A more detailed description of FPGAs can be found in Section 3.3.

As mentioned, the L1T uses data with a reduced resolution to make a rough reconstruction of the event to identify particles such as photons, electrons, muons, and jets. It also calculates global quantities such as H_T and p_T^{miss} ; all of which tend to indicate interesting physics when large. Another important task of the trigger algorithms is the subtraction of pileup, which would otherwise bias measurements such as the momentum of jets. There are multiple trigger algorithms within the L1T: some use the calorimeters and some use the muon detectors. All processed information is then combined in the Global Trigger (GT) which makes the final L1T selection. Events that pass the L1 trigger get sent to the HLT.

Using tracker data is not possible in the current L1T because of the sheer amount of data produced by the tracker. There are over a hundred million individual channels, and each of them would need to be read out at 40 MHz and transmitted to the back-end. This data volume far exceeds the available bandwidth. Additionally, the computational demands of track reconstruction are extremely high, making it impossible for the tracker back-end to perform this task within the tight time constraints of the L1T. However, there is a possibility of using tracker data in the future if CMS increases the L1T latency and/or processing power.

High-level Trigger

Thanks to the L1T, the data rate is reduced to a level where more advanced event reconstruction techniques can be performed. Implemented in software using commercial Central Processing Units (CPUs) located in a counting room at the surface, the HLT reduces the output rate from 100 kHz to 1 kHz. Compared to the L1T, the techniques used at the HLT are closer to what is later

applied in the offline event processing which is the final step before the data can be used for physics analysis. However, the HLT reconstruction techniques are optimised for fast processing. The reconstruction time of one event is of the order of 100 ms and includes the complete raw data from all subdetectors available. If the HLT accepts an event, the events are passed on for storage and a final reconstruction offline.

The HLT runs a sequence of reconstructions and selection steps with increasing complexity. Selections start by using information from the calorimeters and the muon detectors as they require less CPU than processing the data from the tracker. Once the majority of the events have been filtered away, the CPU-expensive track reconstruction can be performed.

The availability of tracking in the HLT increases the momentum and spatial resolution of particles, as well as improves the probability of correctly identifying particles and determining which quark flavour a jet originated from, so-called jet flavour tagging. For example, tracks allow one to decide if an ECAL cluster is from a photon or from an electron. The HLT can make use of simplified versions of particle reconstruction techniques such as Particle Flow (PF) [30] which reconstructs and identifies every particle in an event using all subdetectors. The availability of tracker data also makes it possible to accurately find the position of primary and secondary vertices, a procedure also known as vertexing. Vertexing is important for various HLT processes, for example, identifying jets originating from b quarks (b -tagging) and the PileUp Per Particle Identification (PUPPI) algorithm [31] that mitigates pileup. These techniques are described further in Section 2.3.7.

2.3.7 Offline Event Reconstruction

The offline event reconstruction is predominantly based on the output of the Particle Flow (PF) algorithm [30]. This algorithm uses the information from all CMS subdetectors to identify final-state particles, such as electrons, muons, photons, charged hadrons, and neutral hadrons. It does so by linking tracks with calorimeter clusters and tracks from the muon chambers. The PF elements belonging to an identified particle are removed, so other particles cannot use the same elements. Muons are identified and removed first. To build a muon physics object, a muon detector track is linked to a compatible track from the tracker. Muons are also reconstructed using a tracker track matched with at least one hit in the muon detectors [32]. Charged hadrons are identified from tracker tracks projected to the HCAL and correlated with clusters. A similar procedure is followed for the electrons using the ECAL. The remaining clusters in the HCAL and ECAL are considered to be neutral hadrons and photons, respectively.

Once the particles have been identified, the PF objects can be used for vertexing, pileup removal, calculating the p_T^{miss} and other physics quantities of interest. Pileup can cause incorrect measurements of e.g. the energy and mo-

momentum of the relevant hard-scattering collision. The main method used for pileup mitigation in the CMS offline reconstruction is the PileUp Per Particle Identification (PUPPI) algorithm [31, 33]. The PUPPI algorithm removes all charged particles with tracks not originating from the primary vertex. Neutral particles do not have any tracker information, thus it is difficult to determine if they originated from the primary vertex. Instead, the neutral particles are down-weighted based on the probability that they originate from pileup, where the probability is calculated from the neighbouring non-pileup charged particles.

Another important part of the offline event reconstruction is the reconstruction of jets. The most common technique used by CMS is the anti- k_T (AK) algorithm [34]. The algorithm clusters the event's PF particles in the $\eta - \phi$ plane depending on the relative momentum between the particles. The cluster tends to be circular and centred around the particle with the highest momentum in the jet, where the distance between particles in a jet is defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (2.5)$$

The size of the reconstructed jets can be specified using the distance parameter which corresponds to the largest distance ΔR between any two particles within the jet. The most common types of reconstructed jets have a distance parameter of 0.4 (AK4) and 0.8 (AK8). Once the jets have been reconstructed, various jet flavour tagging techniques can be used to determine what flavour created the jet.

Jet Flavour Tagging

Tagging jets originating from b quarks, b -jets, is very important as many interesting physics signatures, both for SM and BSM physics, have b quarks in their final state [35]. Both Higgs boson and top quark decays include b quarks the majority of the time. When a b or c quark hadronises, the resulting jet includes a hadron containing a quark of the original flavour. Jet flavour tagging techniques use the fact that heavy-flavour hadrons have a long enough lifetime for their decay to be displaced with respect to the primary vertex, resulting in a secondary vertex as illustrated in Figure 2.7. The lifetime of around 1.5 ps for b hadrons allows the hadron to travel centimetres before it decays, depending on its momentum. For c hadrons, the lifetime is about 1 ps. The distances travelled by the heavy-flavoured hadrons are long enough to be measured by the CMS detector.

Because of the mass of the b quark, the decay products of a b hadron tend to have a larger transverse momentum with respect to the jet axis compared to the other constituents of the jet. In other words, the b -jets tend to be wider than other jet flavours. The fact that b -jets include an e or μ 20% of the time is also considered in the b -tagging methods.

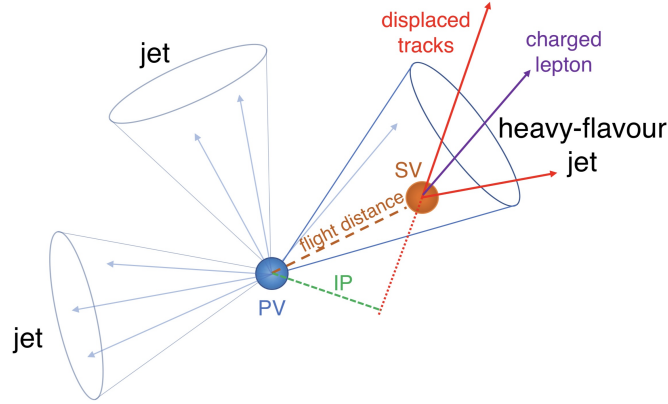


Figure 2.7: An illustration of jets originating from the primary vertex (PV), including a heavy flavour jet with a secondary vertex (SV) from the decay of a b or c hadron [35]. The displacement of the tracks from the secondary vertex with respect to the primary vertex is called the impact parameter (IP).

Sometimes, the decays of heavy neutral particles result in a $b\bar{b}$ pair, e.g. $H \rightarrow b\bar{b}$. If the momentum of the original particle is high, this would lead to a boosted $b\bar{b}$ pair, which appears as a single but wider jet. The hadronisation products from a single quark are typically clustered by an AK4 jet, while a boosted $b\bar{b}$ pair is usually represented by an AK8 jet. Because the corresponding AK8 jet from a double- b jet consists of two b jets, it is possible to use jet substructure techniques to identify subjets, as illustrated in Figure 2.8. To identify a boosted double- b jet, b -tagging methods be applied to the AK8 jet, or to the two subjets separately. Descriptions on two so-called double- b -taggers are found in Section 6.2.

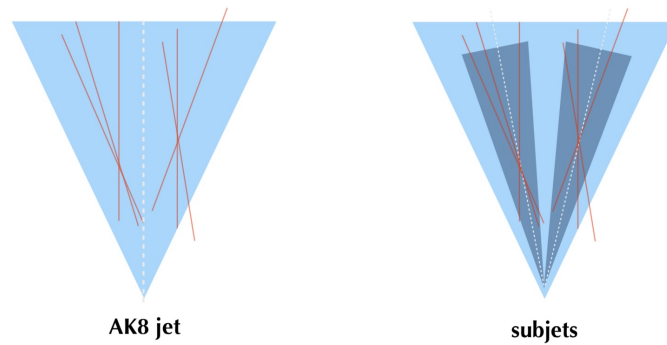


Figure 2.8: An illustration of boosted double- b jets represented as an AK8 jet (left) or as an AK8 jet with two subjets (right) [35].

2.4 High-Luminosity LHC

After over a decade of operation, the LHC is still the most powerful particle accelerator in the world and has enabled invaluable contributions to high-energy physics. Following Run 3, a prolonged shutdown period is scheduled for the installation of the High-Luminosity LHC (HL-LHC) [13] to improve the physics potential at the LHC. The HL-LHC will substantially increase the number of simultaneous proton-proton collisions, thus increasing the instantaneous luminosity and the data rate, with the aim of reaching an integrated luminosity of 4000 fb^{-1} in just a decade. This would increase the integrated luminosity by an order of magnitude, thus opening up the possibilities for many new or improved physics searches. The primary physics goals of the HL-LHC are twofold: to make precision measurements of the SM, particularly involving interactions and self-interaction of the Higgs boson; and to continue the search for BSM physics such as SUSY. These processes are rare, with low cross sections, necessitating a large volume of data to make any conclusions.

The HL-LHC is planned to be operational at the beginning of 2029 [36], also known as Phase-2, and a timeline of the project can be seen in Figure 2.9. The upgrade to the HL-LHC will achieve an instantaneous luminosity of up to $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, with a pileup of 200 proton-proton interactions on average. This is a great improvement compared to the current luminosity of $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. The increase in instantaneous luminosity is mainly accomplished by improving the magnets which focus the beams to increase the probability that collisions will occur. Additionally, an effort to operate the collider at 14 TeV will be made, a limit set by the magnetic fields produced by the magnets which bend the proton trajectories inside the LHC. As the collision energy rises, the number of rare processes typically increases as well, offering a greater opportunity to observe and study these phenomena.

2.5 The CMS Phase-2 Upgrade

The CMS experiment has several major goals for the HL-LHC, including precision measurements of the SM with a focus on the Higgs boson. With an increase in integrated luminosity, the CMS Collaboration hopes to achieve measurements of very rare Higgs decays that would provide more detailed information about the Higgs boson's coupling to other particles, including itself. Additionally, the experiment aims to continue the direct searches for BSM physics. Greater amounts of data will increase the detector's sensitivity to new particles and extend the search for BSM physics to higher masses, such as in the hunt for SUSY. However, the current detector would not be able to cope with the conditions at the HL-LHC.

One of the main challenges at the LHC is radiation damage to various components of the CMS detector. When particles from the proton-proton col-

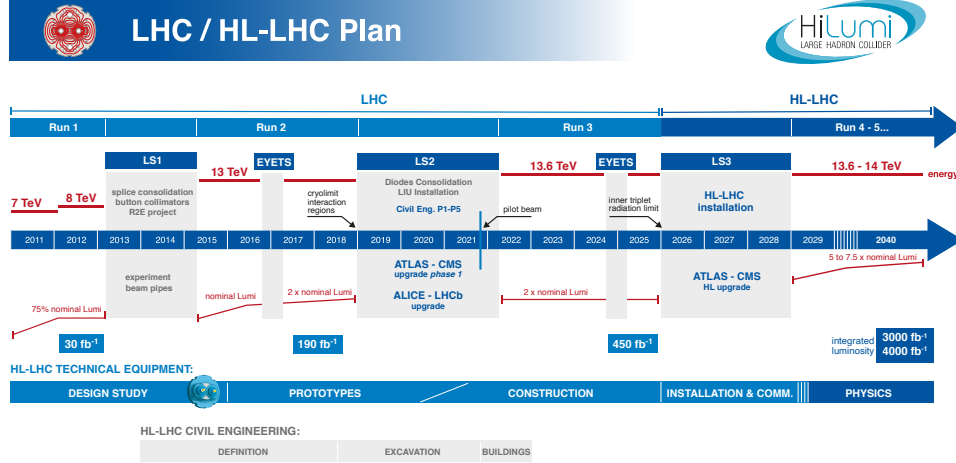


Figure 2.9: A timeline of the LHC and the HL-LHC projects [36]. The timeline includes the data-taking runs, the Extended Year-End Technical Stops (EYETSs) for accelerator and detector maintenance, as well as Long Shutdowns (LSs) for major upgrades.

lisions interact with the detector and the on-detector electronics, the components progressively degrade. Sensors lose their efficiencies, the noise increases, and they might even go completely dead. At the end of Run 3, many of the CMS subsystems will be in need of replacement. The increased luminosity will also pose greater challenges with pileup. With up to 200 collisions per bunch-crossing, the CMS detector must have a high enough granularity to be able to distinguish the hard interactions from the pileup. This is not only important for the offline event reconstruction, but also for the trigger system as this level of pileup will increase the rate of false triggers. Therefore, in view of the HL-LHC upgrade, a major Phase-2 upgrade of the CMS detector [8] will take place.

The current pixel and strip tracker will be completely replaced by the Inner Tracker and the Outer Tracker, respectively [24]. The tracker upgrades will have a higher granularity and increased coverage in the forward region, up to approximately $|\eta| \approx 4$. The Phase-2 tracker layout can be seen in Figure 2.10. The Inner Tracker will be located closest to the beamline and consist of two types of pixel modules: a small module with space for two readout chips placed in the inner-most layers, and a larger module with space for more pixels with four readout chips. The pixels will have a size of $25 \times 100 \mu\text{m}^2$ or $50 \times 50 \mu\text{m}^2$, which is smaller than for the Phase-1 pixel detector. In total, there will be 3900 pixel modules corresponding to around 2 billion readout channels. The sensor modules and the readout electronics will both be radiation resistant to

cope with the high luminosity and light to reduce the particle scattering for better resolution. However, the Inner Tracker will still need to replace parts a few times during Phase-2, hence it has been designed such that it can easily be inserted and extracted. The Outer Tracker will be equipped with sensor modules with on-detector electronics that can filter pairs of hits depending on the p_T . These hit-pairs will be used for fast track reconstruction for the L1T and allow algorithms such as PF and PUPPI to be used in the L1T for an improved trigger performance. The front-end readout chips also include deeper buffers than in the current detector, which allows for a longer latency of the L1T system, and larger bandwidths for an increased readout rate. Further details on the Outer Tracker and the L1 tracking are found in Section 2.5.1 and Chapter 3, respectively.

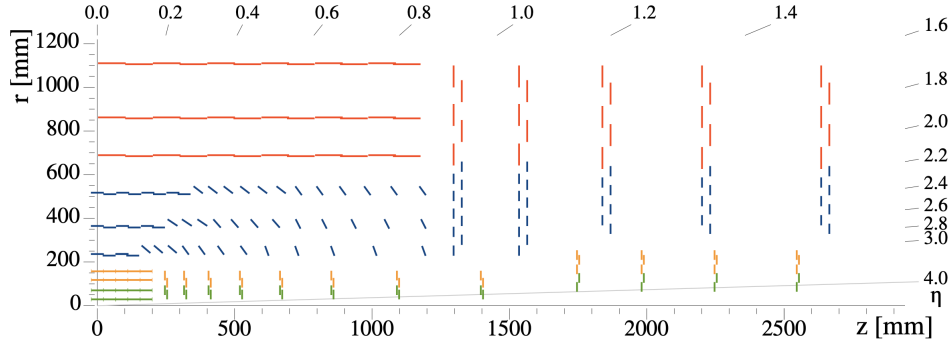


Figure 2.10: The tracker layout for the HL-LHC [8]. It shows one-quarter of the tracker in the $r - z$ plane, with the beam along the z -axis and the interaction point at the origin. The Inner Tracker is depicted in green and orange, with the smaller sensor modules in green. The Outer Tracker PS and 2S sensor modules are depicted in blue and red, respectively.

A new timing detector made of very fast silicon sensors, the MIP Timing Detector (MTD) [37], will be installed between the tracker and the calorimeters. The MTD will be thin to not obstruct the performance of the ECAL, and measure charged particles with a time resolution of around 30 ps at the beginning of its lifetime, degrading to 60 ps at the end. The 200 nearly simultaneous collisions within a bunch-crossing occur with a Gaussian spread of around 5 cm along the beamline and with a 180-200 ps time span. Therefore, the MTD will be able to differentiate two collisions even if the corresponding particles originated from the same location. This will lead to an increased event reconstruction performance and pileup rejection.

Both the ECAL and the HCAL will need upgrades, however, much of the barrel layers is in good condition and meet the requirements of taking up to 4500 fb^{-1} of data [38]. The barrel layers are expected to handle the increased instantaneous luminosity after replacing the back-end electronics of both sub-

detectors and the front-end electronics of the ECAL. As a consequence of the significant radiation damage in the forward regions, the ECAL and the HCAL endcaps will be completely replaced by the High Granularity Calorimeter (HGCAL) [39]. Layers of silicon sensors and absorber materials, such as lead and copper, will make up the electromagnetic section of the new endcap, as well as the hadronic section closest to the beamline. These regions will endure high occupancy and radiation levels, hence silicon sensors were chosen because of their proven radiation hardness. Absorber layers made of stainless steel and plastic scintillators read out by silicon photomultipliers will make up the hadronic section in the less harsh regions. The HGCAL will have a high granularity and also be able to give timing information with a precision of around 50 ps, both of which are important to assign the correct particle shower to the correct primary vertex.

Much of the muon chambers will be kept as is, but the electronics will be replaced because of the increased readout frequency and latency requirements at the HL-LHC. Additional muon chambers [40] will also be installed for increased coverage in the forward region and for keeping the L1 trigger rate low. Additional muon chambers included in the L1T yield more accurate measurements of p_T , which in turn reduces the trigger rate which is dominated by mismeasurements of the momentum. The CMS Phase-2 upgrade also involves maintenance and improvements of the CMS back-end, such as upgrades of the L1T [29] and the HLT [41]. To maintain both a viable trigger rate and physics sensitivity, major changes to the L1T are needed. The trigger needs to efficiently select rare events to be able to reach the physics goals of performing high-precision measurements of the SM, e.g. within the Higgs sector, and searches for BSM physics. Simply increasing the thresholds of the current trigger quantities would reduce the physics potential. As previously mentioned, the L1T will include data that has not previously been available, such as tracker information. The detector coverage and granularity have also increased, which the new trigger will profit from. To allow for more data being processed and for more complex techniques to keep the rate at a manageable level, the back-end electronics of the L1T and the processing units of the HLT will be upgraded. Furthermore, the trigger latency for the L1T (HLT) will be longer at 12.5 μ s (~ 1 s) thanks to the deeper readout chip buffers to give time for more complex algorithms to be run. The readout bandwidth has been increased for a trigger rate of 750 kHz (7.5 kHz) thanks to improved readout chips.

2.5.1 Outer Tracker

The original strip tracker was designed to withstand radiation for 500 fb^{-1} of integrated luminosity, which is close to the value expected after Run 3. Consequently, a completely new Outer Tracker [24] is planned for the HL-LHC upgrade that can reliably function up to 4000 fb^{-1} . Because of the great

difficulties that an increased pileup brings, the detector will be designed such that tracker information can be included in the L1 triggering for the first time at CMS. With the availability of tracker tracks and the increased L1T latency, it will be possible for the L1T to make use of more advanced event reconstruction techniques such as PF and PUPPI. This will improve particle identification and pileup rejection. Tracks will also make the resolution of the p_T better, which will result in better resolution for various quantities such as p_T^{miss} and H_T . Tracks also allow for primary vertex reconstruction, and it might even be possible to reconstruct in real time displaced tracks, i.e. tracks not originating from the beamline due to long-lived particles which tend to indicate interesting physics. Without tracks in the L1T, the trigger rate could be as high as 4000 kHz, way past the maximum limit of 750 kHz.

The Outer Tracker will also have an increased granularity to keep the occupancy low (1% or less) and improve track separation. Compared to the previous tracker, the new Outer Tracker will contain less material, which leads to better tracking as well as improved ECAL resolution. The Outer Tracker back-end will also be upgraded and is split into two systems: the Data Trigger and Control (DTC) and the Track Finder (TF). The DTC will communicate with the front-end, but also distribute tracker information to the TF which will process the Outer Tracker data for the L1T.

Front-end

Like its predecessor, the Outer Tracker will feature a barrel with multiple cylindrical layers. At each end of the barrel, endcaps consisting of disk-shaped layers enclose the Outer Tracker volume (Figure 2.10). The Outer Tracker will consist of modules with two closely separated silicon sensor layers mounted on top of each other, as shown in Figure 2.11. There are two types of modules: the PS module featuring a macro-pixel and a strip sensor, and the 2S module featuring two strip sensors parallel to each other. The three barrel layers closest to the beam will consist of PS modules, as will the inner half of the disks. The strips are positioned such that the resolution is high in the direction that the charged particles bend, the ϕ -direction, while the macro-pixels enable the module to also have a good spatial resolution in z (r) for the barrel (endcap). For even better resolution and minimisation of the material traversed by the particles, some of the barrel PS modules have been tilted such that the sensor plane is perpendicular to particle trajectories originating from the interaction point as seen in Figure 2.10. The three outer layers of the barrel, and the outer half of the disks, will consist of 2S modules, as it is not necessary to keep the high z -granularity where the occupancy is lower.

The sensor modules also include front-end ASICs that correlate the signals from the two adjacent sensors to differentiate between low-momentum and high-momentum particles, as illustrated in Figure 2.11. Low-momentum particles tend to move at a larger angle compared to the direction perpendic-

ular to the sensor plane (Equation 2.3), causing hits in the two sensor module layers to be farther apart. The front-end ASIC is configured with a window in the sensor further away from the beam such that only pairs of nearby hits, referred to as a stub, get transmitted to the back-end for further processing. The window is chosen such that the stubs correspond to a charged particle with a transverse momentum greater than 2 GeV. This filtering reduces the data rate by a factor of 10-20, and around 15 000 stubs get passed to the tracker back-end, which prepares the data for the L1T. While the stubs are being processed by the back-end, the full events including the low momentum hits which were not included as stubs are stored on buffers on the front-end ASICs waiting for an event selection.

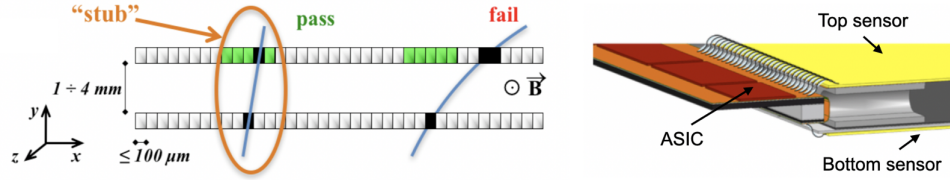


Figure 2.11: The stub-finding concept (left), and a depiction of a sensor module with two closely separated sensors (right) [8]. A charged particle traverses both sensor planes. The particle will bend more if it has low momentum, thus a stub window, shown in green, in the top sensor can be chosen such that only particles compatible with $p_T > 2 \text{ GeV}$ get transmitted by the on-detector electronics.

Back-end

The Outer Tracker back-end consists of two layers of hardware systems. The first system is the Data Trigger and Control (DTC) which can control and communicate with the Outer Tracker front-end. For every event, the DTC distributes the stubs to the second hardware layer, the Track Finder (TF). The TF processes the stubs to form tracks that are transmitted to the L1T. Both systems also transmit monitoring data to the CMS DAQ system. The DTC also distributes the L1A from the L1T to the front-ends, which forwards the full raw data from the front-end to the DAQ system. A diagram of the data-flow is shown in Figure 2.12.

To handle the high data rates, the system is split into nine equally sized regions in ϕ (40° each) that operate in parallel. These regions, or sectors, are determined by how the cables from the Outer Tracker front-end are routed out of the detector. For each sector, there are 24 DTC boards working in parallel, each communicating with a designated region of the Outer Tracker. As the two types of sensor modules are slightly different, a board only needs to communicate with one type of module. In total, there will be 216 ($9 \times$

are running the track-finding algorithm in parallel but for different events. This is done by components which shuffle the links between the DTCs and the TFs every bunch-crossing. Therefore, a board receives a new event every 18 bunch-crossings, i.e. every time-multiplex period. There will be 164 (9×18) boards in total, excluding any potential spares. The six links interfacing each TF board to the L1T systems are operating at a rate of 25 Gb/s.

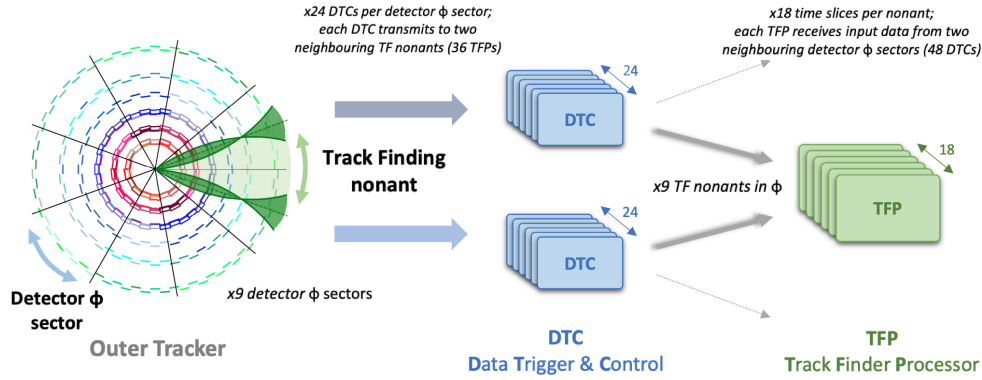


Figure 2.13: The detector sectors and a track-finding sector (nonant) in green, in an $x - y$ view of the Outer Tracker [29]. The overlap region is marked in dark green. The track-finding sectors are shifted by 20° relative to the detector sector edge, hence only two detector nonants and its DTCs need to be involved for each Track Finder (TF) sector.

The tracker back-end boards will follow the Advanced Telecom Computer Architecture (ATCA) [44] industry standard for communications equipment. The standard includes specifications for the processing boards as well as the shelves for the boards that provide cooling and power. The architecture is designed to be flexible, scalable, and fault tolerant. If a subsystem encounters failures, the rest of the system can still be kept running.

Serenity board

The DTC system will run on Serenity boards [45], which are custom back-end boards developed for the CMS upgrade to handle the high data rates at the HL-LHC. Serenity is a general-purpose board compatible with most subdetectors for the Phase-2 upgrade. It has been designed to be flexible allowing use by many subdetectors in various capacities. Serenity is still under development, but one Serenity board prototype can be seen in Figure 2.14. The most recent board consists of one FPGA (Xilinx Virtex UltraScale+ VU13P [46]) that will implement the data processing. Previously (as shown in Figure 2.14), there was space for two FPGAs, but this was changed as no subdetector required more than one FPGA. The boards also contain a wide range of connectivity op-

tions (both remote and direct physical connections), a small System-on-Chip (SoC) with a Linux operating system to control and manage the board, and high-speed connectors for fast communication and data transmission between back-end systems. The Serenity board includes six bi-directional connectors to transmit or receive data, with the ability to take either electrical or optical transceivers with up to twelve channels. There are also electrical connections between the two FPGAs in the old Serenity versions. For the Outer Tracker, which has a high data throughput, 12-channel optical transceivers operating at 25 Gb/s will be used to connect the DTC to the TF.

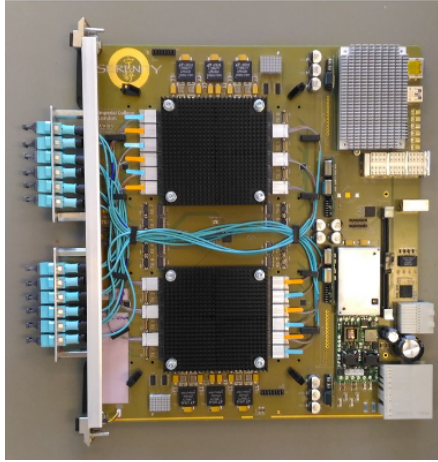


Figure 2.14: An older version of the Serenity board with two FPGAs under the black heatsinks. The optical transceivers are seen with turquoise and orange labels. Optical fibres and connectors are seen in turquoise.

Apollo board

The TF system consists of custom boards called Apollo [47, 48], which can be used in various LHC experiment systems for high-speed readout and triggering. The Apollo boards are under development and several prototypes have been made. Apollo consists of a service module and a command module. The service module handles communications with external ATCA service systems and delivers the clock signal and power to the rest of the board. Depending on the need of the subsystems using the board, the command module can accommodate multiple FPGAs that will perform the data processing, along with connectors for optical links up to 25 Gb/s. In total, there are more than 100 channels available. There are also 25 Gb/s electrical connections between the FPGAs. To control and monitor the Apollo board itself, a System-on-Chip (SoC) with Linux is used. The TF boards will have two FPGAs (Xilinx Virtex UltraScale+ VU13P) installed. Apollo will also be used for the Inner Tracker DTC system.

Chapter 3

L1 Tracking Algorithm

3.1 Introduction

To maintain the physics sensitivity and reduce the impact of pileup at the HL-LHC, the CMS Collaboration will introduce a radical improvement to the L1 trigger system. This involves the use of a novel L1 Track Finder (TF) that utilises tracker information obtained from the stubs of the Outer Tracker, as discussed in Section 2.5.1. The goal is to reconstruct all tracks with a transverse momentum greater than 2 GeV within 4 μ s. For each event with 200 pileup vertices, which corresponds to around 200 tracks, of the order of 10 000 stubs will be produced as a consequence of effects such as particle scattering and high sensor occupancies in the innermost layers. The tracks will be reconstructed by a track-finding algorithm implemented on FPGAs and then sent to the L1 trigger where they will be used in algorithms that perform vertex finding and particle identification. The availability of tracks will, for example, improve p_T measurements, jet reconstruction, pileup rejection, as well as improved isolation of particles such as electrons and muons.

Various approaches to track-finding exist, but to be considered effective, they must reconstruct nearly all tracks with sufficient resolution for accurate p_T measurements and vertexing. At the same time, the algorithm must avoid reconstructing any fake or duplicate tracks as this could compromise the L1T performance. The track-finding algorithm must also be fast to meet the L1T latency requirement and robust against any potential sensor inefficiencies. The track-finding algorithm chosen by CMS consists of a road-searching algorithm that connects stubs to create track candidates [49], and a Kalman Filter to do the final track parameter fit [42].

This chapter will provide an overview of the L1 track-finding algorithm and its corresponding firmware implementation. Results from one of the firmware modules will be given, the so-called VMRouter, as well as results from the firmware simulations and demonstrations of slices of the final project.

3.2 Track-Finding Algorithm

The nine TF sectors, described in Section 2.5.1, each run the same track-finding algorithm independently and in parallel. In the road-search (tracklet) portion of the algorithm, illustrated in Figure 3.1, two close-by stubs in a pair of adjacent layers are selected to form a seed. The seed and the beamline are combined to create a helical track estimation or tracklet. Only seeds which create a tracklet with a $p_T > 2$ GeV are considered. The four helix parameters consist of the initial z coordinate (z_0) and azimuthal angle (ϕ_0), the inverse radius ($1/R$), and the tangent of the polar angle ($\tan \theta$). The tracklet is then projected onto the other layers to identify any stubs consistent with the extrapolated helix by comparing the distances between the stub and the tracklet. A potential track candidate has been found if there are four or more layers, including the seed layers, with matching stubs to the track projection. The same algorithm steps are repeated for all seeds. It is important to note that the same stub can be used in multiple seeds and that the seeding process is performed in multiple layer combinations to increase redundancy and improve robustness against inefficient sensor modules.

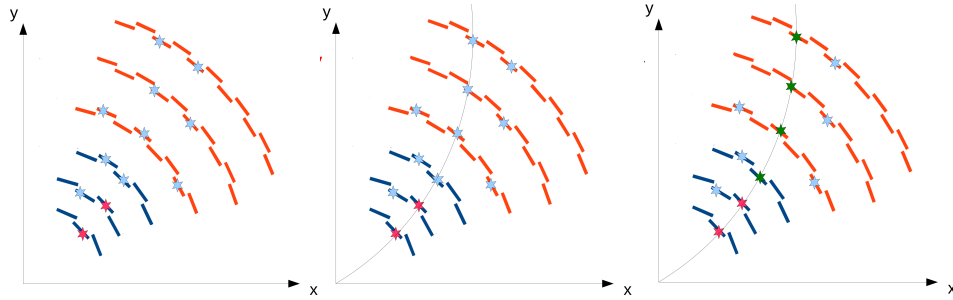


Figure 3.1: The concept of the track-finding algorithm [49]. First, two stubs are paired to create seeds (left). Using the seed and the origin, a helical track is calculated and projected to the other layers (middle). Finally, stubs that are close enough to the track estimation in the other layers are matched to the track (right).

Multiple occurrences of the same track candidate can arise because of the formation of seeds in various layer pairs. To address this, track candidates that share a significant portion of stubs are merged to form a single track.

The final track parameter fitting is accomplished using a Kalman Filter [42], which determines the helix parameters by using all matching stubs in the track for a more precise result, in contrast with the tracklet, where the helix parameters were determined using only the two stubs from the original seed. The Kalman Filter uses a series of measurements, i.e. the stubs assigned to the track, to provide an estimate of the particle trajectory. Stubs are added to the Kalman Filter one by one, with each stub further constraining the estimate of the helix parameters, assuming that the stubs were produced by

the same particle. The Kalman Filter can reject stubs that seem inconsistent with the fitted trajectory, and reject tracks altogether if the resulting track does not have enough stubs. If there are multiple stubs in a layer, due to the track merging, the Kalman Filter will choose the one stub best matched to the track. Other methods of performing the fit of the stubs to estimate the track parameters are also possible, such as a χ^2 fit [49]. However, the implementation of rejecting inconsistent stubs which might appear during the stub matching or track merging steps would complicate the methods.

The quality of the track is assessed using a Boosted Decision Tree (BDT) before the track is sent to the L1T. A BDT is a machine learning algorithm used for classification. Each track is assigned a value for the quality that is appended as an additional track parameter. This extra parameter can be used in the L1T to remove fake tracks and to select tracks that have a better track reconstruction quality.

The performance of the track-finding algorithm has been studied using a software emulation of the algorithm written in C++. The number of bits for each variable and constant in the emulation matches what is implemented in firmware. Using simulated top-quark pair-production events with 200 pileup, as these events are complex and result in many tracks, the algorithm was shown to reconstruct around 95% of the tracks [29], as shown in Figure 3.2. The fraction of fake and duplicate tracks were around 5% and 2.5%, respectively. The resolution of z_0 increases with $|\eta|$ (Figure 3.2) as it is more challenging to determine the origin of a track at large angles. For small angles, the z_0 resolution is about 1 mm, which aligns with the level of precision at which the L1T vertexing algorithm operates. It has been shown, using emulations of both the L1 track-finding algorithm and the L1T, that including tracks in the L1T does improve the L1T performance such that the trigger rate requirements can be met [29]. Chapter 4 explores one study of how including tracks affect the L1T.

3.2.1 Virtual Modules

To make the seeding and stub matching more time efficient when implemented in hardware, the track-finding algorithm organises the stubs into bins based on layer and ϕ . The ϕ -bins are called Virtual Modules (VMs), and are illustrated in Figure 3.3. When performing the track finding, only bins in the various layers that are compatible with a $p_T > 2$ GeV particle need to be considered. In the seeding procedure, stubs in a given VM in the inner layer are only compatible with stubs in a small subset of the outer layer VMs corresponding to particles of $p_T > 2$ GeV originating from the beamline. When matching stubs to a tracklet projection, only stubs in the VMs close to the projection will be considered for matching. Therefore, the use of VMs simplifies the hardware implementation as well as reduces the processing time and the overall size of

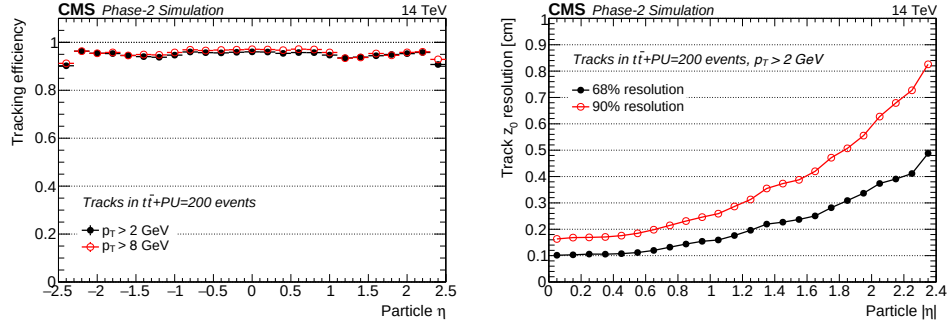


Figure 3.2: The L1 tracking efficiency as a function of η (left) for all particles with $p_T > 2$ GeV ($p_T > 8$ GeV) in black (red), and the track z_0 resolution (right). For each $|\eta|$ bin, the z_0 resolution is acquired from the interval at which 68% (black) and 90% (red) of all tracks with $p_T > 2$ GeV are included. Both plots were made using simulated top-quark pair-production events with 200 pileup [29].

the algorithm as only a fraction of the number of all possible stub combinations needs to be investigated.

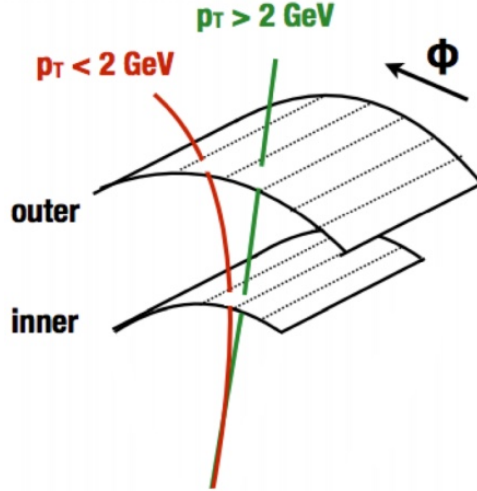


Figure 3.3: The seeding process using VMs represented by the dashed lines [50]. A VM in the inner layer is only compatible with certain outer layer VMs, which would correspond to a track from a particle of $p_T > 2$ GeV (green). Consequently, tracks falling outside these VMs will not be taken into account (red).

3.3 Field Programmable Gate Arrays

Field Programmable Gate Arrays [51] are integrated circuits that can be programmed to perform specific tasks according to user requirements. Their high speed and easy reconfiguration make them useful for a wide range of applications. FPGAs consist of numerous Configurable Logic Blocks (CLBs) that can be interconnected to create custom logic circuits and perform complex computations. Because the CLBs operate independently, FPGAs are very effective at parallel processing. However, to ensure synchronization, FPGAs rely on a clock which can be specified by the user, and only a certain number of tasks for each CLB can be performed within one clock cycle. Operations which exceed one clock cycle fail to meet the timing requirements, causing the implementation to become unsynchronised.

The decision to use FPGAs for the L1 Track Finder was based on their speed and commercial availability, as they offer a cheaper and more accessible alternative to developing expensive Application-Specific Integrated Systems (ASICs). An FPGA allows for flexibility as the algorithm can be modified at any time, something that would not be possible with an ASIC. Developing an algorithm on ASICs would also require expert knowledge in ASIC design while implementing an algorithm on an FPGA is more accessible to the general particle physicist. Using CPUs, as for the HLT, would not be possible as they do not operate at fixed latencies, which is required for the L1T. The track-finding algorithm will be implemented on Apollo boards, each with two Xilinx Virtex Ultrascale+ VU13P [46] FPGAs installed. Newer generations of FPGAs exist, however, either the specifications do not meet the project requirements and/or they have not been available on the market for very long, thus their performance has not been studied. During some integration tests, a smaller Xilinx Virtex Ultrascale+ VU7P was used.

FPGAs have a fixed number of resources available, but the type and quantity of resources vary depending on the model and manufacturer. This description focuses on FPGAs from Xilinx. The CLBs consist of other components such as Flip-Flops, Look-Up Tables (LUTs), and multiplexers. A Flip-Flop is a component that can store a single binary value and is used to save the current state between clock cycles. LUTs offer fast outputs of logic operations by storing a predefined list of outputs for every combination of inputs, thus complex calculations can be avoided. Multiplexers select one of several input signals as the output signal.

Random-Access Memories (RAMs) and Digital Signal Processors (DSPs) are additional components of an FPGA. There are two types of RAMs in the Xilinx Virtex Ultrascale+ family: the smaller and more power-efficient Block RAM (BRAM) and the larger, faster, but more power-consuming UltraRAM (URAM). DSPs are microprocessors specialised for quickly performing mathematical operations such as multiplication. Additionally, FPGAs have a limit on the number of available high-speed input and output ports (I/Os), which

is determined by the number of gigabit transceivers (called GTY). The specifications of the VU7P and VU13P FPGAs can be seen in Table 3.1.

	VU7P	VU13P
CLB Flip-Flops	1 576 320	3 456 000
CLB LUTs	788 160	1 728 000
Max. Distributed RAM (Mb)	24.1	48.3
Block RAM Blocks	1 440	2 688
Block RAM (Mb)	50.6	94.5
UltraRAM Blocks	640	1 280
UltraRAM (Mb)	180.0	360.0
DSP Slices	4 560	12 288
GTY Transceivers 32.75 Gb/s	80	128

Table 3.1: Specifications of two models of Xilinx Virtex Ultrascale+ [46].

3.3.1 Vivado and Vivado HLS

Configuration of the FPGA is done in a programming language, generically known as firmware. Firmware is written in a Hardware Description Language (HDL), which is used to describe digital logic circuits. While multiple languages are available, one of the most common is VHDL [52]. To enable software developers to design firmware without having to learn an HDL, a software engineer approach based on the so-called High-Level Synthesis (HLS) [53] is available. This tool allows users to describe firmware using high-level programming languages, such as C++ [54], which are automatically converted into an HDL. The tool attempts to perform this conversion in a way that meets timing requirements, as well as minimising the resource usage within the FPGA.

The two main software packages used for writing firmware for the L1 tracking are Vivado and Vivado HLS from Xilinx. In Vivado, the user can write firmware using an HDL and run simulations of the design. The design is then synthesised, i.e. converted into a low-level representation of the digital circuit in terms of its logical components and their interconnections. In the implementation stage, the synthesised design is mapped to the specified FPGA to generate a physical layout of the circuit. This step is also referred to as place and route. Typically, the resources and connections between the resources are placed and routed in a way that minimises congestion. Congestion refers to the situation when there are too many resources and interconnects that are competing for a limited number of available routing resources, which can lead to delays and low performance or even non-convergence of the synthesis. Using various optimisation techniques, the user can specify directives to optimise the

design, also called floorplanning. Once the placing and routing are finished, a bitstream file is created for programming the FPGA with the firmware design.

Vivado HLS is a tool for designing firmware using HLS. The firmware is written in C++ but with additional compiler directives, called pragmas. All C++ memory allocation of the program occurs at compile-time as the code represents actual physical logic elements on the FPGA which cannot be changed unless the FPGA is reprogrammed. This makes some C++ functionalities unavailable inside HLS as allocating memory dynamically at runtime is not possible. The pragmas allow the user to control the synthesis process to guide the firmware implementation. For example, there is a pragma, called unroll, to implement each for-loop iteration in parallel for faster processing. Unrolling a loop will reduce the latency, i.e. the number of clock cycles it takes to process the data. However, this will increase resource usage compared to running something sequentially, which would reuse the same resources for each loop iteration. To save resources while keeping the latency low, a pipeline pragma can be specified along with the number of clock cycles (a minimum of one clock cycle) between the start of each loop iteration. Pipelining breaks up the operation within a loop into a series of smaller sub-operations. Each sub-operation takes the specified number of clock cycles, if possible, thus multiple loop iterations can be processed at the same time but with a slight delay as the same resources are used for each iteration.

In Vivado HLS there are two different simulations: one that runs the C++ as is, and one that simulates the design after synthesis. The synthesised design can then be implemented and exported as an Intellectual Property (IP) core. An IP core is a pre-designed block of firmware, which can be reused and easily integrated into an FPGA by simply specifying the input and output ports.

3.4 Track-Finding Firmware Implementation

The track-finding implementation in firmware is split into several parts, called processing modules, and connected as a chain. A simplified version of the track-finding chain can be seen in Figure 3.4, and a description of each processing module is given in Section 3.4.1. The system is time-multiplexed and new input will arrive every 18 bunch-crossings, thus each processing module has a fixed latency of 18 bunch-crossings as well. Each processing module handles one event at a time and a majority of the modules write the output to intermediate memories, where the data is stored temporarily. Data not processed within the fixed latency will be discarded. The memories are read by the processing modules downstream, which was processing the previous bunch-crossing, thus all modules are operating at all times. Some algorithm steps also produce data for modules much further downstream in the chain, thus they will be stored in the memories for longer before being read. The modules at the end of the chain, which do the filtering and fitting of the tracks,

stream their output directly to the next module. No intermediate memories are necessary for these modules as the tracks can be processed as they arrive, thus the filtering and fitting modules combined only take one time-multiplex period. The total latency of the algorithm is then mainly determined by the time-multiplex period multiplied by the number of processing modules which write data from an intermediate memory, plus one time-multiplex period for the filtering and fitting. With the chain illustrated in Figure 3.4 and a minimum bunch-crossing time of 25 ns, the total latency of the algorithm would be 3.6 μ s ($18 \times 8 \times 25$ ns), which is less than the maximum limit of 4 μ s.

To increase the efficiency of the track-finding chain, multiple copies of each processing module and intermediate memory operate in parallel, resulting in faster data processing and reduced processing time. This is done by letting each copy process distinct data based on e.g. ϕ -regions/VMs or seed combinations. An illustration of a track-finding chain is provided in Figure 3.5. The intermediate memories correspond to the VMs, making it simpler for downstream modules as it ensures that modules only receive the data relevant to their specific VMs, reducing the amount of data that needs to be processed. Copies of memories are created if two processing modules need to access the same data, to avoid conflicts due to limitations on read and write operations per clock cycle. The detector coverage of the VMs depends on the average number of stubs in the bin, thus they are different for each layer and module. Small VMs would lead to fewer stubs to process for each module but increase the likelihood of a processing module being idle because of data unavailability. Idle modules imply inefficient usage of the resources on the FPGA.

In order to maximise the data throughput while keeping the resource usage low, the processing modules are designed to commence processing one stub or track candidate every clock cycle. In other words, the entire process is pipelined. Each processing module is responsible for processing one bunch-crossing at a time. The minimum time between bunch-crossings is 25 ns and with a time-multiplexing period of 18, the processing module will have 450 ns (18×25 ns) to process a bunch-crossing. The clock is set to 240 MHz which means that a maximum of 108 (450 ns $\times$ 240 MHz) stubs or tracks can be processed by each module. Any remaining unprocessed stubs will be discarded. The impact of data loss was shown to be negligible as a consequence of the redundancy when creating seeds from multiple layer combinations [49].

The processing modules are implemented either in VHDL or using HLS and exported as IP cores. The intermediate memory modules are implemented in VHDL. A script generates the VHDL top-level function that connects all modules to create the final track-finding chain. VHDL is typically used when precise control over resource usage is required and/or when the implementation in an HDL is relatively straightforward. Apart from the fact that HLS makes firmware designing available for non-experts, it also makes it easier to implement, simulate, and verify complex operations that may be hard to write directly in an HDL.

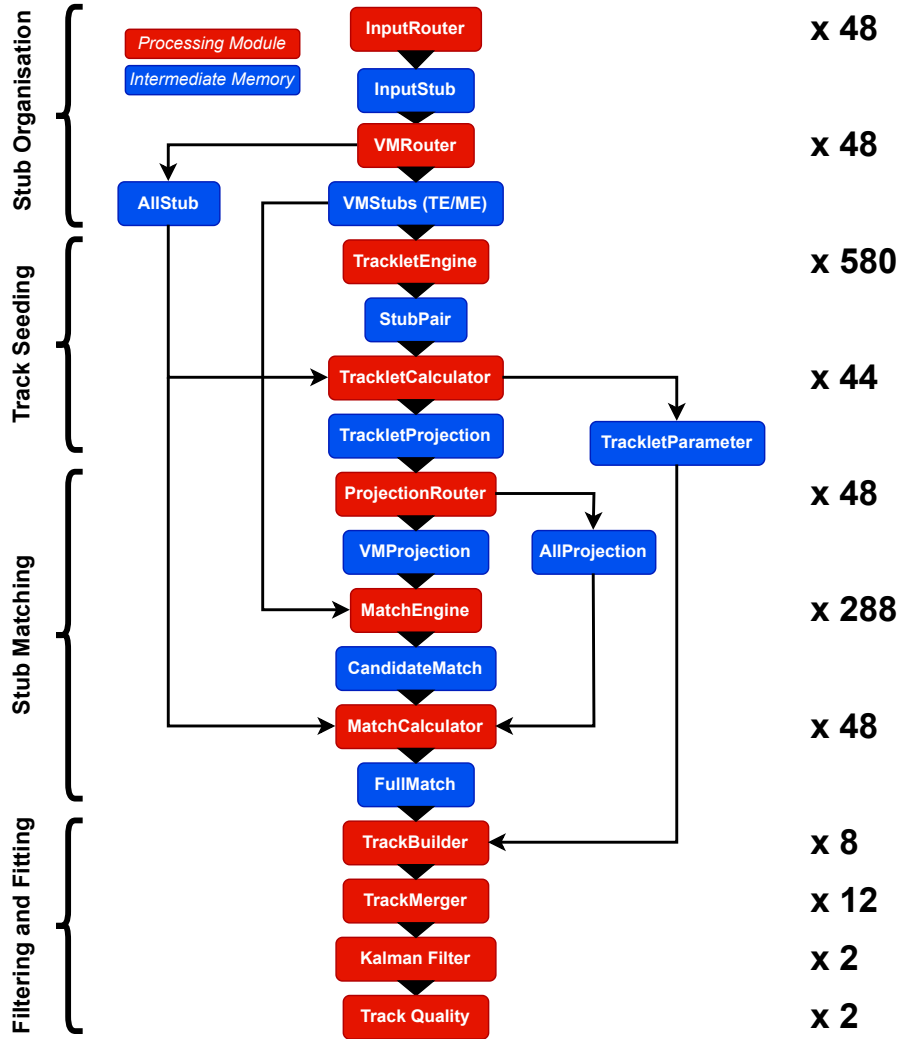


Figure 3.4: A simplified flowchart of the track-finding algorithm with processing modules in red and intermediate memories in blue. Section 3.4.1 provides a summary of the purpose of each module. The number of instances of each processing module per sector is also indicated.

3.4.1 Processing Modules

In this subsection, the processing modules of the track-finding algorithm will be presented in sequential order, along with brief explanations of their functions. With the exception of the Kalman Filter, which was coded in VHDL, all processing modules were created using HLS.

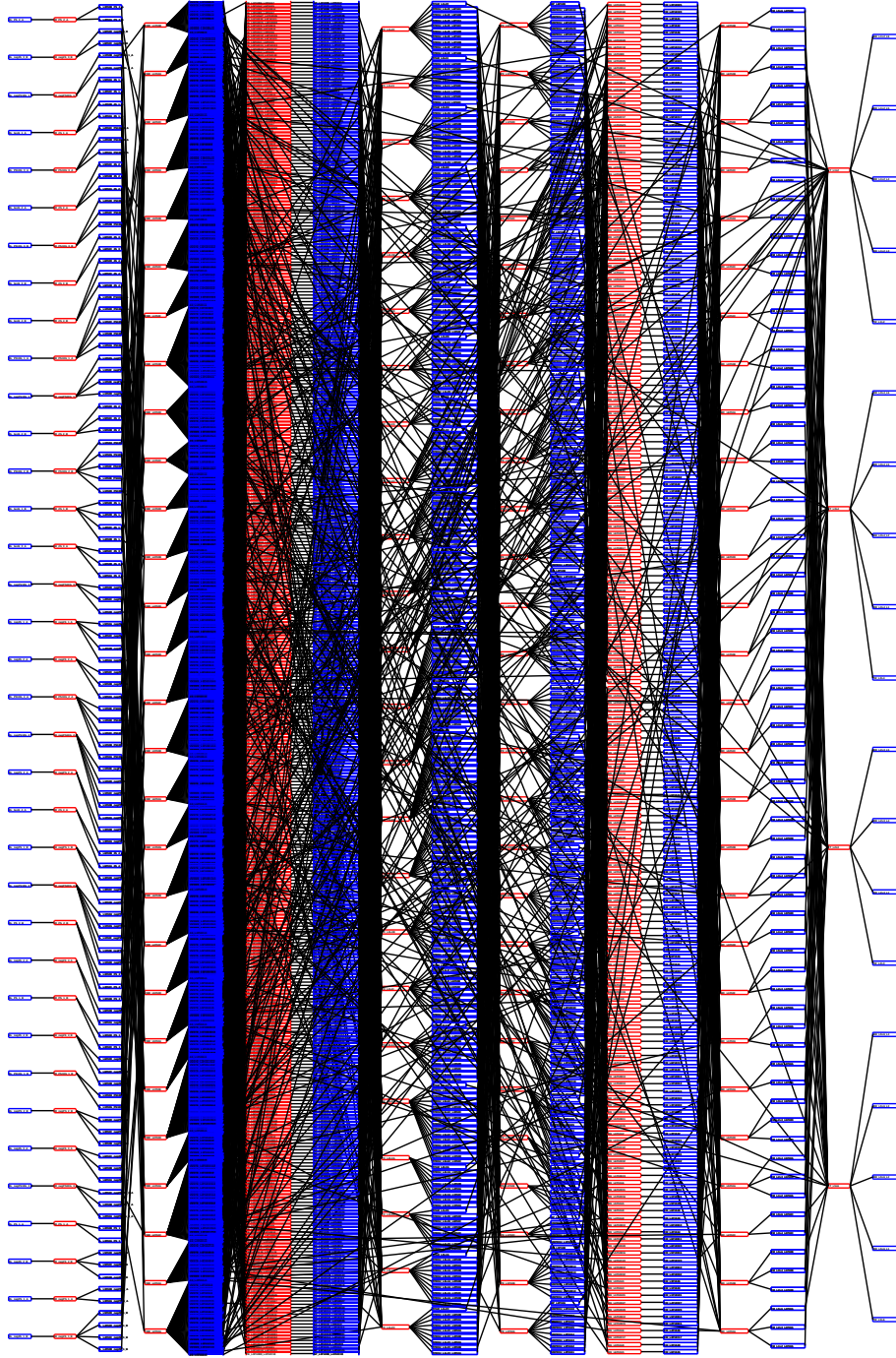


Figure 3.5: An overview of the tracklet algorithm using only modules for the tracker barrel. All processing modules up to the Track Builder module were included (including copies). This corresponds to approximately two-thirds of the full project. Processing modules and intermediate memories are depicted in red and blue, respectively. Stubs are fed in from the left, and tracks are outputted on the right.

Input Router

Upon arrival from the DTC, the stubs are not sorted in any meaningful manner. The first module, known as the Input Router (IR), organises the data according to which tracker layer it originated from. To aid the downstream module, the Input Router also sorts the stubs into large regions in ϕ (four regions for each layer, except the first barrel layer which has eight). The resulting output is stored in intermediate InputStub memories.

Virtual Module Router

The VMRouter (VMR) reads the InputStub memories and organises stubs into VMs. Each VM corresponds to an intermediate memory, with multiple copies created to avoid conflicts when read by multiple processing modules. The VMR saves stubs for multiple downstream modules: the AllStub memories contain the full stub data, while the data in the VMStub memories is shortened based on the needs of the processing module that will read from that specific memory to save resources. The VMStub memories also contain the address of the corresponding stub in the AllStub memory such that the full information can be retrieved further down the chain. Further details on the VMR are given in Section 3.5.

Tracklet Engine

The Tracklet Engine (TE) identifies stub pairs that could correspond to particle tracks with a $p_T > 2 \text{ GeV}$. It benefits from the use of VMs as only a few VMs need to be checked. The stub pairs, or seeds, are formed in several barrel (L) and disk (D) layers: L1L2, L2L3, L3L4, L5L6, L1D1, L2D1, D1D2, and D3D4. This redundancy compensates for sensor inefficiencies as particles travel through the tracker. The seeds are saved in StubPair memories.

Tracklet Calculator

In the Tracklet Calculator (TC), each seed from the StubPair memories along with the beamline is used to create a tracklet. The StubPair memory only contains the reference to which stubs made up the seed, thus this address is used to look up the full stub information in the AllStub memories. The TC then uses the tracklet helix parameters to project it to other layers in the detector. The tracklet parameters and projections are saved in the TrackletParameter and TrackletProjection memories, respectively.

Projection Router

The Projection Router (PR) sorts the tracklet projections based on VMs. The sorted output projections are saved inside two memory types: the VMProjec-

tion and the AllProjection. The latter intermediate memory includes more bits per projection as required by downstream modules.

Match Engine

Stubs are matched to the projected tracklet in all layers that were not already used to form the seed. The matching is performed by the Match Engine (ME) using the projections from the VMProjection memories and stubs from the VMStub memories. The matched stubs are then stored in the CandidateMatch memory.

Match Calculator

The Match Calculator (MC) calculates the distance between each matched stub stored in the CandidateMatch memory and the tracklet projection in the AllProjection memory. The resulting residuals are then used to decide whether the matched stub should be retained or discarded. The selection criteria depend on which layer or disk the matched stub originated from and which layer and disk combination the tracklet was created from. The matched stubs are saved in the FullMatch memory.

Track Builder

The Track Builder (TB) takes all matched stubs from the FullMatch memory along with the corresponding tracklet parameters and compiles them into unified track objects. The TB also removes any tracks with stubs in fewer than four layers. The selected tracks are then directly streamed to the next module.

Track Merger

As previously mentioned, it is possible to find the same track multiple times because of the redundancy in the seeding. The Track Merger (TM) compares the track objects and merges them if it identifies them as duplicates. The remaining merged tracks are streamed to the Kalman Filter.

Kalman Filter

The previous steps of the algorithm relied on a seed, i.e. two stubs, and the beamline position to determine the track parameters. However, the Kalman Filter uses the beamline and all matched stubs in all layers to perform the final track fit.

Track Quality

The Track Quality module is an implementation of a BDT. Useful quantities, such as the χ^2 of the Kalman Filter fit in $r - \phi$ and $r - z$, are calculated in an additional output step after the Kalman Filter and used as input to the BDT. The BDT evaluates the input from the streamed tracks and calculated quantities, and assigns a quality value to each track based on its likelihood to be genuine.

3.4.2 Intermediate Memories

All intermediate memories in the track-finding chain are implemented with one of two memory templates written in VHDL: binned and unbinned memories. Unbinned memories save the output in the order it was processed, while binned memories can sort the output based on an additional parameter such as r - or z -coordinate. The binned memories are used by the VMStub memories. All data within the memory is stored as a bit vector, where the length of the vector can be specified depending on the memory type. It is up to each module to decode which bits correspond to what data.

As the intermediate memories need to store multiple bunch-crossings at the same time, each bunch-crossing is written to its own virtual block of memory, also called a page. Only two pages are required if the memory is immediately read by the next module in the processing chain: one for writing the output from the upstream module and one for the downstream module to read the previous bunch-crossing's data. However, if the memory is required by a later module, additional pages are necessary to store the data without overwriting it with new information.

3.4.3 Top-Level VHDL

The purpose of the top-level VHDL is to establish signal connections between processing modules and intermediate memories. In total, there are over 1000 processing modules that need to be connected to over 4000 memories. Because of the extensive number of signals that require declaration, a Python script is used to create the top-level in a manner that is simple, scalable, and flexible. This script can take arguments to produce smaller track-finding chains for testing purposes.

The Python script takes input in the form of a list that outlines the interconnections between processing modules and memories. It generates a top-level VHDL which instantiates all relevant HLS IP cores and VHDL memories, in addition to other necessary utilities such as size definitions of each memory type. Each module and memory consists of input and output ports for data read and write operations. Additionally, they include ports to count the amount of data already saved for each event as well as ports to indicate the modules' intent to read or write data. There are also ports to specify the

memory address from which to read or write the data. To maintain synchronisation, each module and memory also includes the clock signal. All these extra ports between the processing modules and memories are included in the top-level VHDL script.

The script also produces a VHDL testbench for simulating the specified chain. The testbench reads input data from text files and stores it in the intermediate memory located before the first module in the chain unless the first module is the IR. As the IR gets its data streamed directly to the module, a special entity which acts like a First In First Out (FIFO) was created. The FIFO-like entity reads the input data one at a time and sends one stub per clock cycle in the same sequential order as it was read. The testbench also writes the content of each intermediate and final memory to output text files to validate the behaviour of the chain.

3.5 VMRouter

The VMR organises the stubs in ϕ and outputs the data in memories corresponding to the VMs. There are in total 48 VMRs in the final chain, each processing a different layer and ϕ -region (four ϕ -regions per layer except for the first barrel layer which is split into eight ϕ -regions). The data is also truncated depending on the information needed by the downstream processing modules that read from the memories. The VMR takes several InputStub memories as input and loops over the stubs one by one until 108 stubs have been processed, or there are no stubs left in the InputStub memories. Each stub is saved in at least one memory based on which layer and ϕ -region is processed. The complete stub data is always saved in the AllStub memory. The stub contains the r -, z -, and ϕ -coordinates, as well as bits for the bend and the variable α . The bend is an encoded value of the track curvature inside the sensor modules, determined by the front-end, and is used for deciding which seeds can be created from the stub. Because of poor radial resolution in the 2S disk modules which can affect the ϕ -measurements, α is an additional variable that can be used to make corrections to the ϕ -coordinate. The bit assignment of the InputStub and AllStub memories is listed in Table 3.2. The AllStub memories contain all stubs that were processed by the VMR. Note that the disk layers consist of both PS and 2S modules and that the total number of bits for a stub within the memories need to be the same because of firmware constraints. That is why the r quantity is padded with three zeros for 2S disk modules.

The VMR also writes stubs to VMStub memories. There are three versions of this memory: the ones read by the Match Engine (VMStubME), and the ones read by the Tracklet Engine (VMStubTE) as the inner or outer layer of the seed. Within the ϕ -range covered by the VMR, there are multiple memories of each type representing different regions in ϕ , i.e. the VMs. There-

Table 3.2: Number of bits for each quantity in the InputStub and AllStub memories.

Quantity	Layer 1-3 (PS)	Layer 4-6 (2S)	Disk PS	Disk 2S
r	7	7	12	7 ^a
z	12	8	7	7
ϕ	14	17	14	14
α	-	-	-	4
bend	3	4	3	4
TOTAL	36	36	36	36

^a ‘000’ + 4 bits

fore, the VMR will look at the ϕ -coordinate and determine in which VMStub memory to save the stub. Not all information is needed by the downstream modules, hence unnecessary information from the input stubs is dropped in the VMStub memories, for example, the quantity α . As the processing modules only process stubs or tracks within a certain region and layer, sometimes only the coordinate bits within that region are needed. The redundant bits which correspond to the region of the VM are dropped, and the remaining coordinate bits we call fine ϕ , fine r , and fine z . Generally, the full precision of these fine-quantities is not necessary either, so the quantities are truncated depending on the precision required by the downstream modules. The radial location of the sensor modules in a barrel layer varies slightly, so to simplify the algorithm, the average radius of the layer is used for calculations. When saving the fine ϕ in the barrel, an extrapolated value of the ϕ -coordinate at the average radius is used. Additionally, an index to the memory position of the corresponding AllStub is saved along with the stub information, such that the full information can be accessed if needed. The stub content for the VMStub memories can be found in Tables 3.3-3.5.

Table 3.3: Number of bits for each quantity in the VMStub memories read by the ME.

Quantity	Layer 1-3 (PS)	Layer 4-6 (2S)	Disk PS	Disk 2S
index	7	7	7	7
bend	3	4	4 ^b	4
fine ϕ	3	3	3	3
fine z/r	3	3	3	3
TOTAL	16	17	17	17

^b ‘0’ + 3 bits

A flowchart of the VMR design can be seen in Figure 3.6. Multiple copies can exist for each memory, one for each module that needs to read the memory.

Table 3.4: Number of bits for each quantity in the VMStub memories read by the TE for inner layer seeding. Overlap refers to the seeding between a barrel layer and a disk.

Quantity	Layer 1-3 (PS)	Layer 5 (2S)	Disk 1 & 3	Layer 1 Overlap	Layer 2 Overlap
index	7	7	7	7	7
bend	3	4	3	3	3
fine ϕ	2	2	2	2	2
z/r	10	10	10	10	10
TOTAL	22	23	22	22	22

Table 3.5: Number of bits for each quantity in the VMStub memories read by the TE for outer layer seeding.

Quantity	Layer 2 & 3 (PS)	Layer 4 & 6 (2S)	Disk 1 & 2 & 4
index	7	7	7
bend	3	4	3
fine ϕ	3	3	3
fine z/r	3	3	3
TOTAL	16	17	16

The AllStub memories written by a single VMR are exact copies of each other, while the stubs in the VMStub memories read by the TE might differ depending on the stub bend. Each TE only creates seeds from one specific inner and outer layer VM combination, thus only stubs with a bend in a consistent direction to the VM combination will be saved to the corresponding VMStub memories.

3.5.1 Resource Usage and Timing

To demonstrate the results of the resource usage and timing of the VMRouter, a single module was used, denoted VMR_L2PHIB: the first few letters of the module name indicate the processing module type (in this case, the VMRouter). The next part specifies which layer or disk the module belongs to, and the last part indicates the phi-region named alphabetically starting from A. There is one VMRouter per phi-region, with up to eight phi-regions per layer. The VMR_L2PHIB module is used as an example as it has the most input and output memories connected to it since the second layer is used for three different seeding combinations. Every input and output require additional logic on the FPGA to implement them, therefore, VMR_L2PHIB can be considered as the worst-case scenario. The number of inputs and outputs for VMR_L2PHIB is specified in Table 3.6. The number of VMs varies depending on

the memory type, as some downstream modules that read from the memories require finer regions. This leads to some memory types having a greater number of “unique” memories, each corresponding to a VM, than others. Additionally, some memories have multiple copies as they are read by multiple modules at the same time, and only one read can be performed per clock cycle.

Table 3.6: The input and output memory types connected to the VMR_L2PHIB module. The number of unique memories of each type is specified, along with the number of copies, and the total number of memories. The InputStub is an input memory, and the rest are output memories.

Memory type	No. unique memories	No. copies	Total
InputStub	4	-	4
AllStub	1	10	10
VMStubME	8	-	8
VMStubTE Inner	4	3	12
VMStubTE Outer	8	5	40
VMStubTE Overlap	2	4	8

When synthesising the module, the HLS compiler provides an estimate of the resource usage and timing. However, these values are unreliable and not used here. Instead, a more precise result is obtained after the firmware is implemented and exported as an IP core. The project satisfies the condition of processing one stub every clock cycle, and the timing estimate for one clock cycle of the module post-implementation is 3.714 ns, thus meeting the requirement to be below 4.167 ns (240 MHz). The post-implementation resource usage for the specified VMRouter is shown in Table 3.7, along with an extrapolation of those values to estimate the total number of resources for all 48 VMRouters. This estimate can be viewed as the upper limit of the resource usage. On a single Xilinx Virtex Ultrascale+ VU13P, the VMRouter uses less than 5% of the total VU13P resources. However, two FPGAs will be used in the final project.

Table 3.7: The resource usage for the VMR_L2PHIB module, with the fraction of resources available on a single Xilinx Virtex Ultrascale+ VU13P. There is also an extrapolation of the resources to estimate the resource usage for all 48 VMRouters in the final track-finding chain. The values do not include intermediate memories.

	VMR_L2PHIB	Extrapolated	Available (VU13P)
LUT	1646 (0.1%)	79008 (4.8%)	1728000
FF	3182 (0.1%)	152736 (4.4%)	3456000
DSP	0 (0.0%)	0 (0.0%)	12288
BRAM	5 (0.1%)	240 (4.5%)	5376
URAM	0 (0.0%)	0 (0.0%)	1280

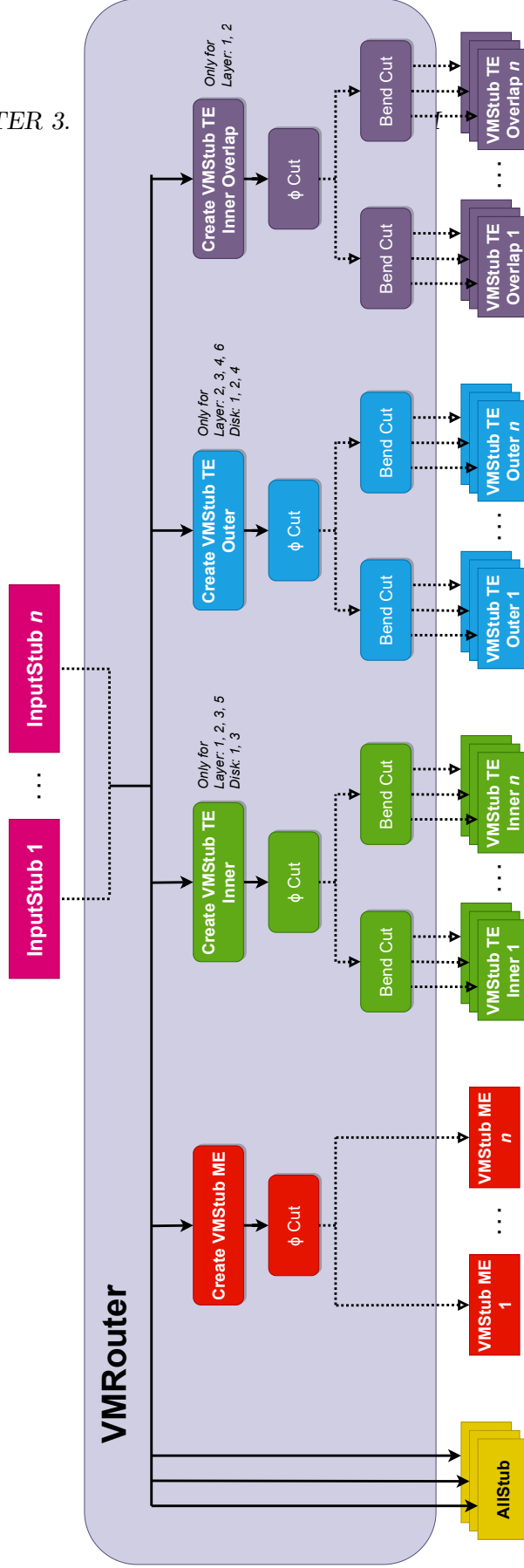


Figure 3.6: A flowchart of the VMRouter design. Every clock cycle, a new input stub is read by the VMRouter. For each memory type, except the AllStub, a function creates the corresponding stub from the input stub and saves it into various memories depending on some cuts. The dotted lines indicate that only some paths are followed for each stub.

3.6 Simulations and Demonstrations

Firmware simulations are performed for each processing module separately before creating longer track-finding chains. Input and output data created by the software emulation of the L1 tracking algorithm are used to confirm the correct operation of each module. This emulation includes all processing modules and intermediate memories. The data stored in the emulation memories is saved and used as input for the firmware simulations, as well as for comparisons between the emulation output and the simulation output. A complete agreement is expected as the emulation uses the same number of bits as the firmware for each variable. The test data are created using up to 1000 simulated top-quark pair-production events with 200 pileup. For testing individual processing modules, the input from the emulation data is stored in the intermediate memories upstream of the module tested. The module under test then reads from those memories, processes the data, and saves it to the downstream memories. The data saved in the output memories are later read and compared to the emulation output. When a chain of modules is tested, the firmware also saves the content of each intermediate memory for easier debugging.

All processing modules have been implemented in firmware and have successfully passed individual firmware simulations, i.e. there was a complete agreement between the output of the firmware and the software emulation. These modules also meet the timing requirements and can process new stub data every clock cycle using a 240 MHz FPGA clock. Short chains, comprising only a few processing steps in the track-finding algorithm, with a small number of modules in each step, have been tested in both simulations and actual hardware. However, further testing of longer and larger chains is described below.

3.6.1 Skinny Chain

The initial synthesis attempt of a long chain that includes modules from the Input Router to the Track Builder was known as the “skinny chain”, depicted in Figure 3.7. This was a reduced version of the complete track-finding chain that centres around a single Tracklet Calculator, and only modules and memories connected upstream and downstream of it were included. The selected Tracklet Calculator only processed seeds formed between barrel layers 1 and 2. The chain encompassed roughly 4% of the modules in the entire project.

The chain underwent initial testing via firmware simulations, with input stubs sourced from the software emulation and streamed into the Input Routers. The resulting output from the Track Builder was recorded and subsequently compared against the emulation. The skinny chain successfully passed the firmware simulations, meeting the timing constraints, and was con-

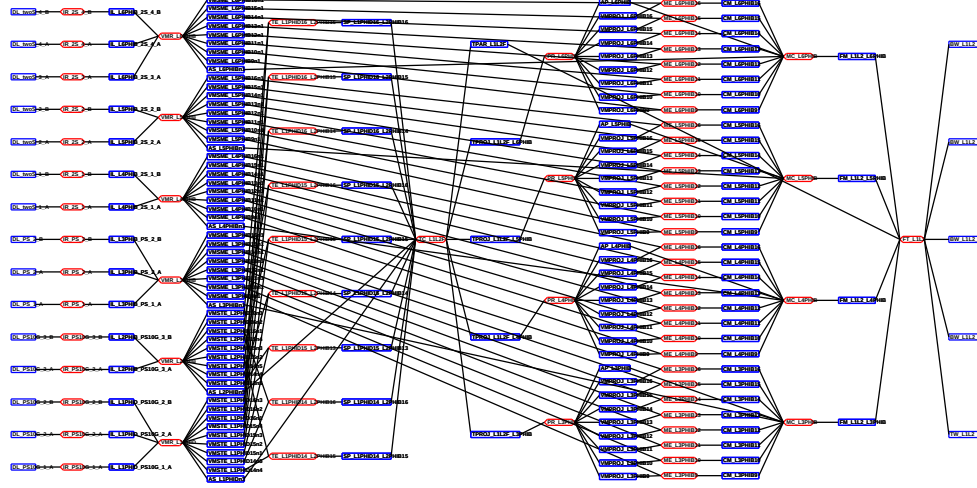


Figure 3.7: An overview of the reduced configuration of the tracklet algorithm, the skinny chain. Processing modules and intermediate memories are depicted in red and blue, respectively. Stubs are fed in from the left, and tracks are outputted on the right.

sequently tested on an FPGA (Xilinx Virtex Ultrascale+ VU7P). A floorplan of what resources within the FPGA were used can be seen in Figure 3.8.

3.6.2 Barrel-Only Chain

Testing of a chain comprising most of the modules in the Outer Tracker barrel layers, excluding the Track Merger, Kalman Filter, and Track Quality modules, has also been conducted. This barrel-only project constitutes about two-thirds of the full project and was set to run on a single Xilinx Virtex Ultrascale+ VU13P during synthesis. However, because of congestion on the FPGA, the project does not meet timing at 240 MHz. When a significant portion of the FPGA's resources is being used, the routing between components can become too long to keep up with the clock rate. As the final track-finding chain is intended to be split over two VU13Ps, ample headroom will be available as this will double the available resources for the project. Progress has been made towards meeting the timing requirements through improved floor-planning of the FPGA resources, and resource usage optimisation is underway. The preliminary resource usage numbers can be seen in Table 3.8.

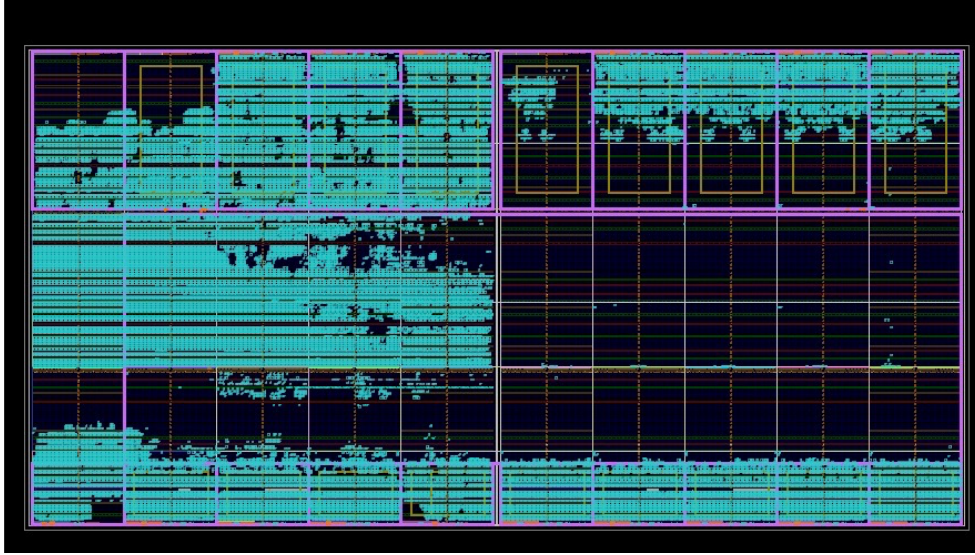


Figure 3.8: The floorplan of the skinny chain on a Xilinx Virtex Ultrascale+ VU7P. Each turquoise section corresponds to a component within the FPGA that is being used by the project.

Table 3.8: Preliminary resource usage for the barrel-only chain along with the available resources on a Xilinx Virtex Ultrascale+ VU13P [55]. The fraction of the resource utilisation of the barrel-only chain on the VU13P is also shown.

	LUT	FF	DSP	BRAM	URAM
Total	711720	1278914	1176	2732	224
Available (VU13P)	1728000	3456000	12288	5376	1280
Utilisation (%)	41.2	37.0	9.6	50.8	17.5

3.7 Combined Modules

There are several improvements that can be made to the track-finding chain architecture, beyond optimising resource usage and floorplanning of the FPGA. To decrease the total latency of the track-finding chain, one approach is to reduce the number of processing steps since each module takes one time-multiplex period to process one event. Decreasing the number of algorithm steps can be done by combining some of the processing modules into larger but more advanced combined modules. The Combined Module chain, depicted in Figure 3.9, implements the same algorithm concept but needs fewer processing modules and eliminates the need for some of the intermediate memories, thereby reducing the total resource usage.

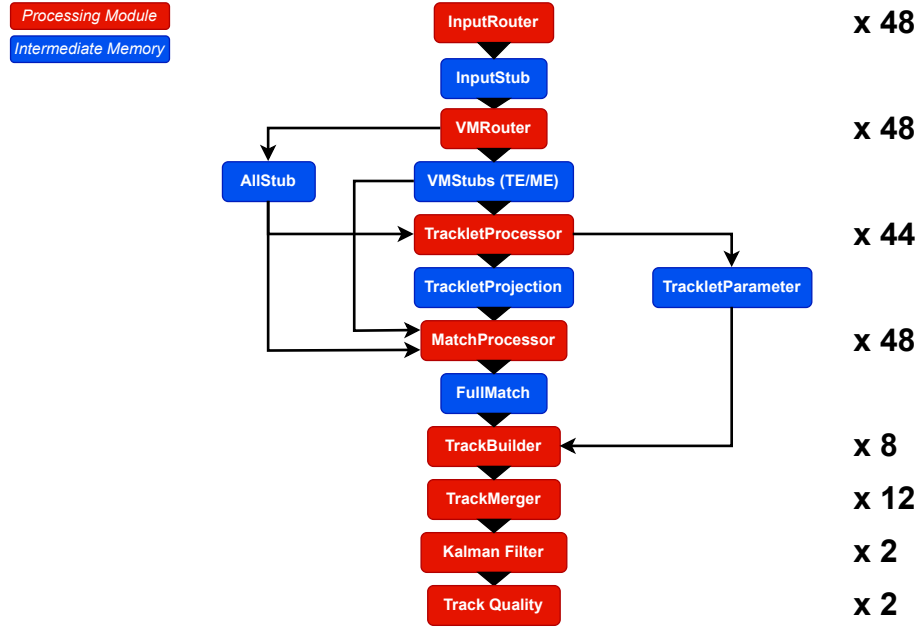


Figure 3.9: A simplified flowchart of the Combined Module version of the track-finding algorithm. Processing modules in red and intermediate memories in blue. The number of instances of each processing module per sector is also indicated.

The Tracklet Engine and the Tracklet Calculator are combined into a single module called the Tracklet Processor (TP). The Projection Router, the Match Engine, and the Match Calculator are combined into the Match Processor (MP) module. This strategy reduces five algorithm steps to two and eliminates the need for three intermediate memory types. Both the Tracklet and Match Processors have been implemented in firmware, as well as a modified version of the VMRouter. The remaining processing modules remain the same. The combined modules are currently used to test the track-finding chain in place of the original modules.

3.7.1 VMRouter for Combined Modules

As some of the modules that read from the memories written by the VMRouter have been combined and modified, changes to the VMRouter were also necessary. The changes included additional bits to the AllStub memories that are read by the Tracklet Processor. These extra bits consist of the index and the fine ϕ bits of the corrected ϕ -coordinate. The stubs saved in these special AllStub memories, the so-called AllStub Inner memories, also need to pass certain cuts in r and z depending on which Tracklet Processor is going to read from the

memory, hence replacing the VMStubTE memories for the inner layers. The content of the AllStub Inner memories can be seen in Table 3.9.

Table 3.9: Number of bits for each quantity in the AllStub Inner memory read by the Tracklet Processor.

Quantity	Layer 1-3 (PS)	Layer 4-6 (2S)	Disk PS	Disk 2S
r	7	7	12	7 ^c
z	12	8	7	7
ϕ	14	17	14	14
α	-	-	-	4
bend	3	4	3	4
index	7	7	7	7
fine ϕ	8	8	8	8
TOTAL:	51	51	51	51

^c ‘000’ + 4 bits

To demonstrate the timing and resource usage of the VMRouter for combined modules, the VMR.L2PHIB module is used again. In Table 3.10, it can be seen that the total number of memories for the module is much smaller than for the original VMRouter (Table 3.6). There is no longer any need for the special VMStub memories in the overlap region between a barrel layer and a disk layer as the Tracklet Processor will use the AllStub Inner and the VMStubTE memories to create all seeds and projections, including the ones in the overlap region. Any ϕ and bend cuts for the VMStubME and VMStubTE memories have been removed as there are no longer hundreds of Tracklet Engines that only process a small subset of possible seed combinations.

Table 3.10: The input and output memory types connected to the VMR.L2PHIB module for the Combined Module chain. The number of unique memories of each type is specified, along with the number of copies, and the total number of memories. The InputStub is an input memory, and the rest are output memories.

Memory type	No. unique memories	No. copies	Total
InputStub	4	-	4
AllStub	1	4	4
AllStub Inner	1	6	6
VMStubME	1	-	1
VMStubTE Outer	1	3	3

The VMR.L2PHIB for the Combined Module chain processes one stub every clock cycle, as required, and the timing estimation for one clock cycle of the module post-implementation is 3.232 ns, i.e. well below the requirement. The

resource usage for the specified VMRouter is shown in Table 3.11, along with an extrapolation of those values to estimate the total number of resources for all 48 VMRouters. The Combined Module version of the VMRouter uses fewer resources as the implementation can be simplified and involves fewer memories compared to the original VMRouter. On a single Xilinx Virtex Ultrascale+ VU13P, the VMRouter uses less than 3% of the total resources.

Table 3.11: The resource usage for the VMR_L2PHIB module for the Combined Module chain, with the fraction of resources available on a single Xilinx Virtex Ultrascale+ VU13P. There is also an extrapolation of the resources to estimate the resource usage for all 48 VMRouters in the final track-finding chain. The values do not include intermediate memories.

	VMR_L2PHIB	Extrapolated	Available (VU13P)
LUT	791 (<0.1%)	37968 (2.2%)	1728000
FF	1123 (<0.1%)	53904 (1.6%)	3456000
DSP	0 (0.0%)	0 (0.0%)	12288
BRAM	2 (<0.1%)	96 (1.8%)	5376
URAM	0 (0.0%)	0 (0.0%)	1280

3.8 Conclusions

The new L1 tracking system at the HL-LHC will enable CMS to reduce the L1 trigger rate to a manageable level while maintaining or even improving physics sensitivity at higher luminosities compared to the current detector. This will be the first time at the LHC that tracker data will be reconstructed and used in the hardware trigger. The L1 tracking has been made possible thanks to the on-module data filtering based on p_T in the new silicon Outer Tracker. In this system, the aim is to reconstruct all charged particles within the Outer Tracker volume with $p_T > 2$ GeV in $4\mu\text{s}$ using an algorithm run on two FPGAs mounted on custom TF boards. The TF is split into nine regions, and each region has its own set of boards without any communication between them. The boards within the same region run the track-finding in parallel but for different events. New events will be received every time-multiplex period of 18 bunch crossings. The track-finding algorithm is split into multiple pipelined steps, called processing modules, which are then connected with intermediate memories to create a chain.

Each of the processing modules has been implemented in firmware and tested individually, meeting the required timing specifications and passing firmware simulations. The simulations were done using input and output data from top-quark pair-production events with 200 pileup. The firmware simulation output is compared against a software emulation of the track-finding algorithm, and a complete agreement is required for the simulation of to pass.

If each individual module would yield similar results when combined to create the full track-finding chain, this would correspond to a track reconstruction efficiency of $\sim 95\%$ which is a sufficient performance for CMS. The fake rate and duplicate rate would be a couple of percent. For one type of processing module, the VMRouter, it is estimated that all 48 VMRouters in the final chain will in total use less than 5% of the resources of one of the two FPGAs that will be used to run the track-finding algorithm.

Scripts have been developed to instantiate and compile all processing modules, and to create the top-level VHDL that connects all the modules to form track-finding chains. The skinny chain, which is a reduced version of the track-finding algorithm and represents 4% of the full project, has been run on hardware. The chain comprises a majority of the different algorithm steps. Additionally, a barrel-only chain has been created, consisting of the majority of modules in the tracker barrel layers, which corresponds to about two-thirds of the final project. Once the barrel-only chain has been tested on hardware, the testing will be extended to the full project. Again, full agreement with the software emulation is expected. To decrease the overall latency and resource usage of the chain, the Combined Module version of the track-finding chain, where some processing modules have been merged, might be used in the future.

Software studies of the L1 track-finding algorithm and the L1T, with the number of bits of each variable and constant matching what is implemented in firmware, showed that the inclusion of L1 tracks does indeed have a positive impact on the L1T performance. In fact, these tracks are essential for effectively reducing the trigger rate while maintaining the desired physics performance.

Chapter 4

Study of the Missing Transverse Momentum Trigger Efficiency

4.1 Introduction

The aim of the L1 Track Finder (as detailed in Chapter 3) is to detect all tracks with a minimum p_T of 2 GeV. These tracks are then passed to the L1T, where a full decision on whether to trigger on the event can be made based on the tracks found and other information from the other CMS subdetectors. A lower track p_T threshold implies more tracks to reconstruct. While having more reconstructed tracks can provide a more complete reconstruction of the event, this comes at the cost of increased FPGA resource usage and bandwidth requirements for both the TF and the L1T. Generally, physics of greater interest is associated with higher track p_T , thus low p_T tracks are less significant such events. However, as pileup at the HL-LHC is expected to increase significantly, effective pileup mitigation becomes essential to properly reconstruct the event. Without removal of pileup at the L1T stage, quantities which are triggered on will be overestimated, thereby leading to an elevated trigger rate exceeding the maximum rate of 750 kHz [29]. For the first time, CMS will use the PF and PUPPI algorithms, both of which are highly dependent on tracks, at the L1T for better particle identification and pileup removal.

With limitations to the bandwidth and resources available, reconstructing all tracks down to a low p_T may not be feasible. For example, changing the minimum p_T requirement from 2 GeV, which the current L1 track-finding algorithm (Section 3) employs, to 3 GeV could reduce the number of tracks by a factor 3 [56]. However, excessively increasing the minimum track p_T requirement could degrade the L1T performance. An alternative track-finding algorithm [42] has been successfully demonstrated in hardware when recon-

structuring tracks with $p_T > 3 \text{ GeV}$, thus an increased minimum p_T requirement could open up more possibilities for the track-finding algorithm implementation. Results from a previous trigger study indicated that a minimum track p_T of 2 GeV instead of 3 GeV would be beneficial for reducing the undesirable background trigger rate [57]. However, this study was based on an offline version of the PF and PUPPI algorithms that used L1 tracks as inputs, which may not match the behaviour of the corresponding online algorithms that the L1T will use. Considering that offline algorithms have been developed for events with excellent track reconstruction efficiency and track resolution, which the offline track reconstruction can deliver, they might be more dependent on tracking information. In this chapter, a study is presented of the relation between the minimum track p_T and L1T efficiency of some trigger quantities and, thus indirectly the resource usage, by using a realistic online L1T emulation.

The L1T triggers on various quantities, such as p_T^{miss} , and if the quantity exceeds the trigger-specific threshold for an event, the event passes the trigger. The p_T^{miss} is calculated as the negative sum of the p_T for all particles in a collision (Equation 2.2). Non-zero p_T^{miss} can occur if some of the produced particles do not interact with the detector, e.g. neutrinos or certain particles in BSM models such as neutralinos (see Sections 1.2-1.3). The p_T^{miss} trigger, therefore, has the potential to detect BSM physics, making it particularly important. It is possible to estimate p_T^{miss} by using muon tracks and the energy deposits in the calorimeters, which is what has been done so far. The L1 tracks can significantly enhance the p_T^{miss} trigger, as PF objects can be used which exhibit a good p_T resolution. Insufficient tracks sent to the L1T due to a high track p_T threshold can result in a degraded p_T^{miss} resolution and the risk of missing signal due to the huge background from pileup. Other trigger quantities such as jet p_T and H_T are also influenced by tracks, but to a lesser extent, and are also studied.

It is also important to study how the minimum track p_T affects the trigger rate on background (minimum bias) events, as the pileup might interfere with the p_T^{miss} calculations, resulting in false triggers. The p_T^{miss} trigger is assigned a maximum allowed bandwidth of 18 kHz [29], and events exceeding this limit will not be forwarded to the HLT. Hence, an inefficient trigger with a high background rate, or a trigger easily affected by pileup, can lead to interesting events being lost.

4.2 Method

The investigation into the impact of minimum track p_T was conducted by running first the software emulation of the L1 TF, and using the TF output as input to the software emulation of the L1T. The TF was not altered for this study. Given that the L1 TF was originally optimised to detect tracks

with a minimum p_T of 2 GeV, reoptimising it for a 3 GeV minimum track p_T would most likely have resulted in an increased tracking efficiency for the new track p_T threshold. However, modifying the TF emulation would have required significant effort, including the retuning of various cuts and potentially redesigning the firmware. The simplification of using the original TF means that the results presented here may exaggerate the L1T performance dependence on the p_T threshold. The output from the TF, i.e. tracks with a $p_T > 2$ GeV, was fed into the L1T emulation, where an additional p_T cut was implemented to simulate the impact on the L1T performance, whilst varying the track p_T cuts. The L1 tracks were only used as input to the PF and vertexing algorithms, though additional algorithms such as PUPPI take the PF objects as input. The algorithms may also have been optimised for a 2 GeV track p_T threshold, but again, the algorithms were left as is. If they had been reoptimised as the p_T threshold was changed, it may have reduced any differences in the performance observed between the 2 GeV and 3 GeV thresholds. However, as the algorithms work for all track $p_T > 2$ GeV, the performance should

Monte Carlo (MC) generated data samples were used for this study. Approximately 100 000 semi-leptonic $t\bar{t}$ events with 200 pileup were used to evaluate the trigger efficiency, as these events tend to be complex and represent a good worst-case scenario. Additionally, around 200 000 minimum bias events with 200 pileup were used to test the trigger's ability to reject uninteresting pileup events. The pileup events, or background, are used to choose which threshold to use for a specific trigger quantity, such as p_T^{miss} , to remain below the upper bandwidth limit.

Normally, the p_T^{miss} trigger covers the region $|\eta| < 3$, as the p_T^{miss} trigger uses data from subdetectors beyond the tracker. The coverage of the Outer Tracker is limited to $|\eta| < 2.4$, so a cut was made to only include this region. Beyond $|\eta| = 2.4$, there is no impact from the L1 tracks, so this region is not studied.

4.3 Results

The background trigger rates are discussed in Section 4.3.1. Results for the impact of minimum track p_T on the p_T^{miss} trigger efficiency are discussed in Section 4.3.2. Other interesting trigger quantities that are directly affected by the availability of L1 tracks, such as the jet p_T and H_T , are discussed in Section 4.3.3.

4.3.1 Background Trigger Rates

The background trigger rates when using a minimum bias MC sample with 200 pileup can be seen for p_T^{miss} in Figure 4.1. The lower the p_T^{miss} threshold, the more frequently the trigger was fired. In the region of the permitted

maximum p_T^{miss} trigger rate at 18 kHz, the difference between 2 GeV and 3 GeV track threshold is in the order of 1%. Choosing a minimum track p_T of 2 GeV versus 3 GeV would have minimal impact on the background trigger rate. A trigger rate of 18 kHz prompts a p_T^{miss} trigger cut of 113 GeV and 114 GeV for minimum track p_T of 2 GeV and 3 GeV, respectively.

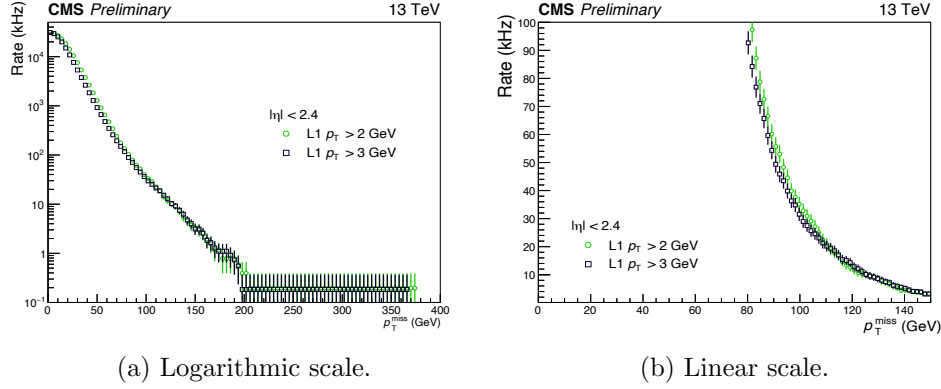


Figure 4.1: The p_T^{miss} trigger rates for different values of minimum track p_T using minimum bias events with 200 pileup. The same plot is illustrated with a logarithmic scale (left) and a linear scale (right). The horizontal axis corresponds to the p_T^{miss} trigger cut.

The background trigger rates for jet p_T and H_T are shown in Figure 4.2. Again, the rate differences between a minimum track p_T of 2 GeV and 3 GeV are small. The maximum L1T trigger rate for jet p_T (H_T) is 70 kHz (11 kHz).

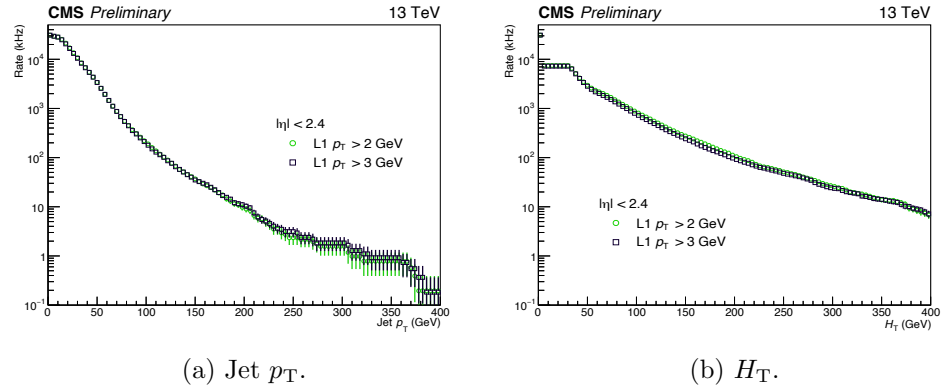


Figure 4.2: The p_T^{miss} trigger rates for different values of minimum track p_T using minimum bias events with 200 pileup. The same plot is illustrated with a logarithmic scale (left) and a linear scale (right). The horizontal axes correspond to the jet p_T and H_T trigger cuts.

4.3.2 Missing Transverse Momentum Trigger Efficiency

Using the p_T^{miss} trigger cuts that yield a rate of 18 kHz, the trigger efficiency was studied. The efficiency is described as the fraction of events in a certain bin of the true p_T^{miss} value that was selected by the trigger. As seen in Figure 4.3, the efficiency when using the original 2 GeV minimum track p_T is slightly higher compared to when using a 3 GeV minimum track p_T . However, the absolute difference between the efficiencies is only up to 5%, though the difference is typically less than that. It can also be seen that the p_T^{miss} trigger efficiency does indeed decrease when increasing the track p_T up to 100 GeV at a constant p_T^{miss} trigger threshold of 103 GeV. The 103 GeV threshold is the default p_T^{miss} threshold for L1T and it is chosen such the trigger rate reaches 95% efficiency at 200 GeV, which is the threshold applied to p_T^{miss} in an offline analysis. A 100 GeV minimum track p_T would suppress almost all tracks going into the L1T. However, the efficiency is still relatively high, as the p_T^{miss} trigger is not only dependent on tracks.

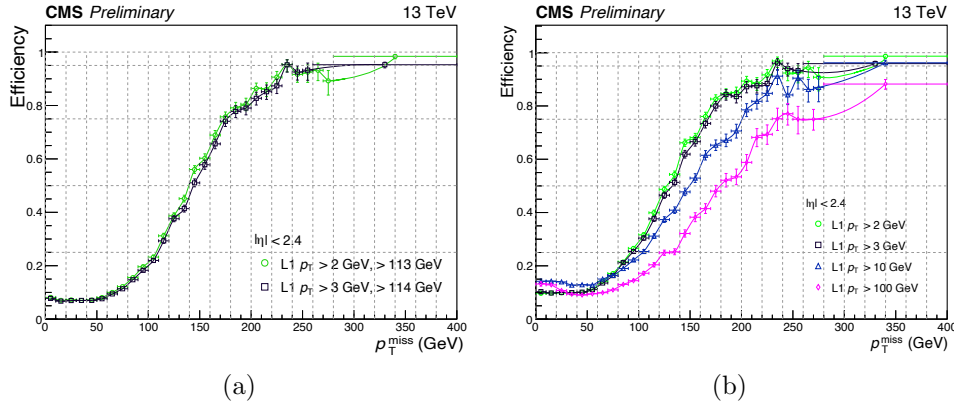


Figure 4.3: The p_T^{miss} trigger efficiency as a function of the true p_T^{miss} derived from the $t\bar{t}$ MC sample. The trigger thresholds in (a) correspond to a constant rate of 18 kHz, whilst (b) use a constant p_T^{miss} threshold of 103 GeV.

In a perfect world, the efficiencies would be a vertical line at the trigger cut, however, because of finite resolution this is not the case. The p_T^{miss} residual between the p_T^{miss} from the L1T emulation and the true p_T^{miss} from the generated MC particles can be seen in Figure 4.4.

4.3.3 Jet p_T and H_T Trigger Efficiencies

The effect of the minimum track p_T was also investigated for trigger quantities such as the jet p_T and H_T , using a constant trigger threshold. The differences between a minimum track p_T of 2 GeV and 3 GeV are very small for the jet p_T , as can be seen for the efficiencies and residuals in Figures 4.5-4.6. The

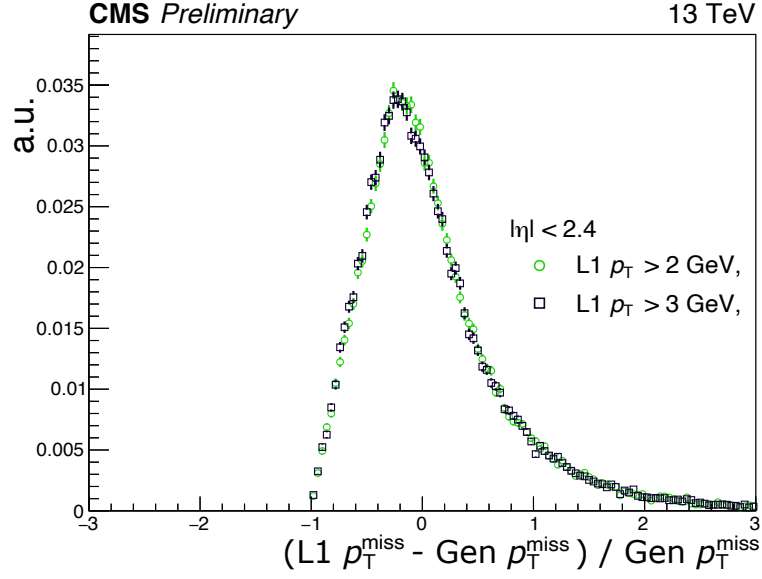


Figure 4.4: The p_T^{miss} residual between the calculated value from the L1T emulation and the true generated value of the $t\bar{t}$ MC sample.

absolute difference in jet p_T efficiency is within 3%, and the scenario with 2 GeV tracks has a slightly better performance.

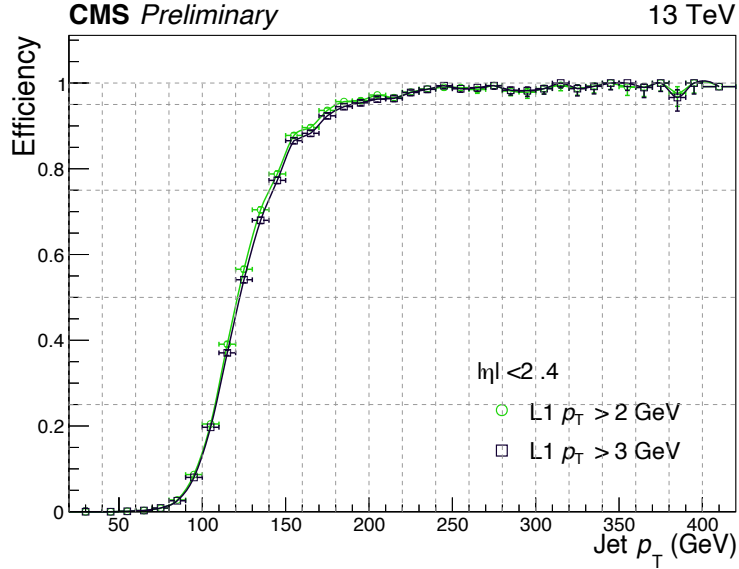


Figure 4.5: The jet p_T trigger efficiency as a function of the true jet p_T derived from the $t\bar{t}$ MC sample. The trigger threshold was set to 121 GeV in both cases.

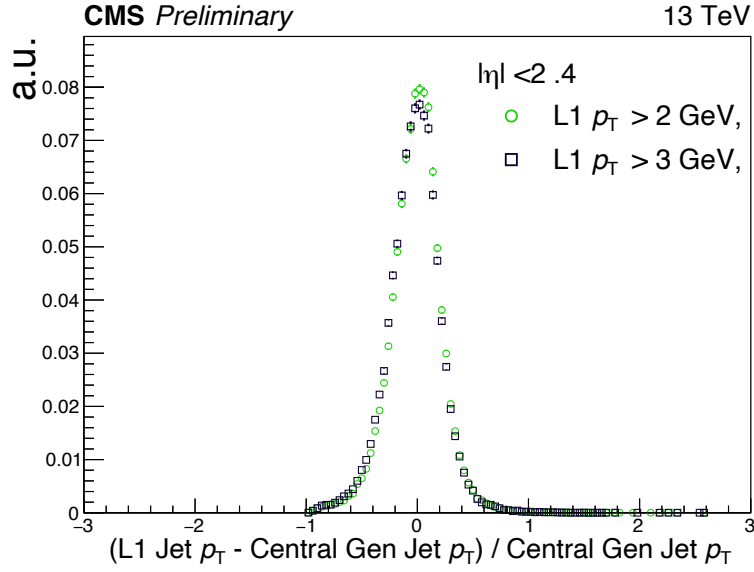


Figure 4.6: The jet p_T residual between the calculated value from the L1T emulation and the true generated value of the $t\bar{t}$ MC sample.

In Figure 4.7, the absolute difference in the efficiency of the H_T trigger is at most 10%. As expected, including more tracks would yield a better trigger. The resolutions, seen in Figure 4.8, are very similar between the two cases.

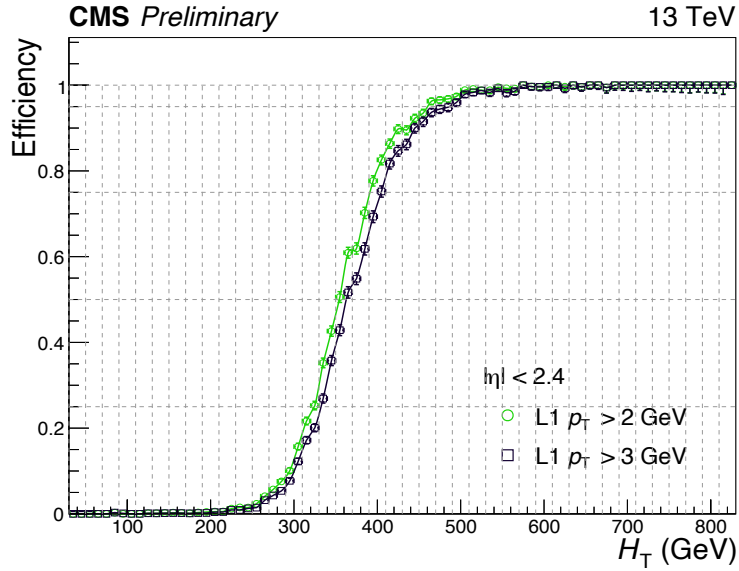


Figure 4.7: The H_T trigger efficiency as a function of the true H_T derived from the $t\bar{t}$ MC sample. The trigger threshold was set to 354 GeV in both cases.

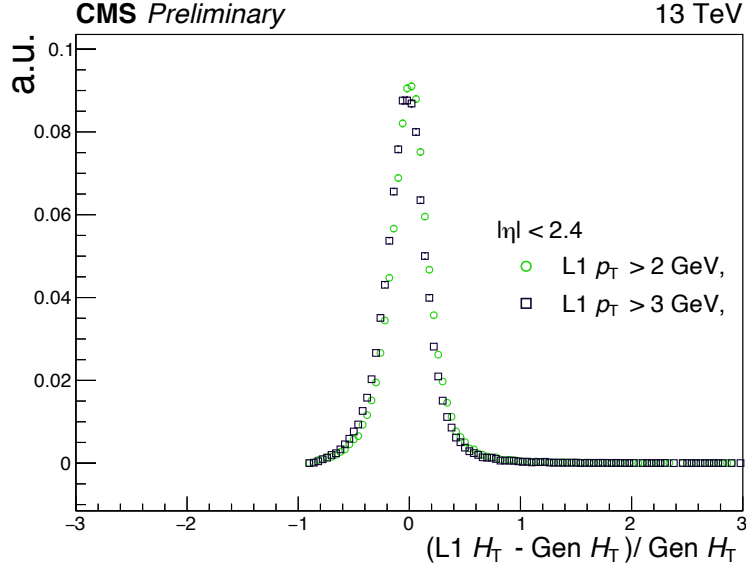


Figure 4.8: The H_T residual between the calculated value from the L1T emulation and the true generated value of the $t\bar{t}$ MC sample.

4.4 Conclusions

This study used an emulation of the L1T to investigate the differences in performance when going from the default minimum track p_T of 2 GeV to 3 GeV. The results show that including more tracks yields marginally better performance for the p_T^{miss} , jet p_T , and H_T triggers. However, the differences are small considering a potential reduction in the number of tracks by a factor of 3. Optimisations of the algorithms to work with 3 GeV could make the observed performance differences even smaller. Going to an even higher minimum track p_T threshold resulted in a non-negligible degradation of efficiency for p_T^{miss} . If the limitations of bandwidth and processing resources in the L1T pose a challenge in the future, adjusting the track p_T could be a potential solution. However, additional investigations into the full L1T are required to ensure the impact of minimum track p_T is small.

Chapter 5

Study of Optical Transceivers for Serenity

5.1 Introduction

The Phase-2 upgrade of the CMS experiment involves improvements across the majority of its systems. With an anticipated surge in data volume, driven by the deployment of higher-granularity subdetectors and an elevated trigger rate, there is a need for enhanced data acquisition and processing capabilities. Consequently, the back-end hardware infrastructure is set to undergo a significant upgrade, equipping it to effectively manage and process data in accordance with the more demanding requirements. Serenity is a custom ATCA board that has been designed for use by multiple subdetectors as part of the CMS back-end upgrade, including the new Outer Tracker DTC system. The board will communicate with the detector, as well as processing and transmitting data to both the L1T and HLT. In the case of the Outer Tracker, the DTC will transmit data to the track-finding boards for L1 track reconstruction. Serenity will employ FPGAs to run detector-specific firmware, while the data transmission will be facilitated by optical transceivers. Details on Serenity can be found in Section 2.5.1. The optical transceivers will convert electrical signals from the FPGA to optical signals for transmission, and vice versa. Currently, the development of the Serenity board is ongoing, and a final decision regarding the selection of optical parts is yet to be made. The choice of transceivers is up to each subdetector, as they have different requirements. This chapter will describe the performance testing carried out on high-speed optical transceivers, including the 12-channel 25 Gb/s Firefly [58] produced by Samtec, which is the main candidate for use in the DTC. Additionally, the results obtained from these tests will be presented.

5.2 Optical Transceivers

To enable communication between Serenity boards and other subsystems, transmitters and receivers are required. Optical fibres are the preferred medium for transmitting data because of their high bandwidth and low signal attenuation over long distances compared to conventional copper wires [59]. As the FPGAs operate using electrical signals, optical transceivers are necessary to convert the electrical signals to optical signals on the transmitter (Tx) path, and vice versa for the receiver (Rx) path. One off-the-shelf option for optical transceivers is the Samtec Firefly [58]. The small size of the Firefly allows it to be placed close to the FPGA on the board, minimising the reduction of the electrical signal quality, or signal integrity, before it reaches the transceivers. Industry-standard connectors (MTP [60]) are used to connect optical fibres to the Firefly transceivers. The transmitters emit light using Vertical-Cavity Surface-Emitting Lasers (VCSELs) [61] at a wavelength of 850 nm. The Firefly transceivers have different specifications depending on the application, with the 12-channel 25 Gb/s version being the main focus of testing in this study. Further information on the tested components can be found in Sections 5.2.1 and 5.2.2.

Clock and Data Recovery (CDR)

High-speed data streams often lack a separate clock signal. In order to receive and interpret the data correctly, the transceiver must extract timing information from the data itself. This is done through a process known as Clock and Data Recovery (CDR), in which the transceiver generates a clock signal from the incoming data transitions. The accuracy of the recovered clock depends on the number of data transitions present in the incoming signal. Once the clock signal is recovered, it is used to synchronise and retime the data stream, resulting in accurate data transmission. The Firefly transceivers are equipped with CDR for both Tx and Rx. The Tx CDR restores the electrical signal after being transmitted from the FPGA to the optical transmitter, while the Rx CDR restores the signal after optical transmission.

5.2.1 Firefly 12-Channel Transceiver

The primary focus of the testing was on 12-channel unidirectional Firefly transceivers with a data rate of 25 Gb/s. The setup consisted of a transmitter-receiver pair, where the Tx and Rx were separate units as the data could only flow in one direction. An example of such a Firefly pair can be seen in Figure 5.1. As the Samtec Firefly transceivers were still in the development phase at the time of the tests, testing was performed on prototype parts. The three earliest versions are referred to as “Alpha”, with “Beta” referring to subsequent prototypes. Mainly the Beta parts will be discussed here.



Figure 5.1: A Firefly transmitter (orange label) and receiver (blue label) pair connected to a Serenity board. When in use, a heat sink is mounted on top of the transceivers to maintain a low operating temperature to avoid damage to the parts. The FPGA is covered by the heat sink in the bottom left corner.

The study involved a total of eight Firefly transceiver pairs. The naming convention for the parts follows a system where the first letters indicate the institute that acquired the parts, followed by the month and year of delivery. If multiple parts were delivered in the same month, an extra letter was added at the end in alphabetical order. Table 5.1 presents a list of the tested parts, along with their corresponding firmware installed by Samtec. All transmitters have the same firmware version, and so do all the receivers.

Earlier tests were carried out on three generations of Alpha parts, which showed poor performance and unreliable behaviour. These parts were delivered to various institutes participating in the Serenity project and were tested on multiple Serenity boards. Results from Alpha parts are available in Appendix B.

5.2.2 Firefly 4+4-Channel Transceiver

In addition to testing the 12-channel Firefly parts, a more mature 25 Gb/s 4+4-channel bidirectional part from Samtec was also tested to provide a basis for comparison. The 4+4-channel part had been tested previously with good results. The bidirectional feature of this part allows for simultaneous data transmission in both directions, thus eight channels in total with 4 channels available in each direction. The 4+4-channel transceiver could serve as a backup alternative to the 12-channel part. However, its implementation could

Table 5.1: The Beta Firefly transceivers that were delivered to Imperial College London in 2022.

Name	Type	Serial Number	Firmware Version
ICjul2022a	Tx Beta	UA220600X9	00120606
ICjul2022a	Rx Beta	UA2136018W	00130C03
ICjul2022b	Tx Beta	UA22060123	00120606
ICjul2022b	Rx Beta	UA21360186	00130C03
ICjul2022c	Tx Beta	UA2206012R	00120606
ICjul2022c	Rx Beta	UA2136017Y	00130C03
ICjul2022d	Tx Beta	UA2206012H	00120606
ICjul2022d	Rx Beta	UA213601AB	00130C03
ICjul2022e	Tx Beta	UA2206012J	00120606
ICjul2022e	Rx Beta	UA213601A6	00130C03
ICjul2022f	Tx Beta	UA2206012Z	00120606
ICjul2022f	Rx Beta	UA2136019Q	00130C03
ICjul2022g	Tx Beta	UA2206011A	00120606
ICjul2022g	Rx Beta	UA213601BK	00130C03
ICjul2022h	Tx Beta	UA2206012M	00120606
ICjul2022h	Rx Beta	UA213601BU	00130C03

require additional connectors on the final Serenity boards to accommodate the increased number of transceivers needed.

5.3 Test Setup

The setup used for testing the Firefly transceivers included a Serenity board with an FPGA from Xilinx, optical fibres, and the transceivers. Most measurements were performed using a Xilinx Virtex Ultrascale+ VU7P FPGA. To create a loopback configuration, the Tx and Rx for each Firefly transceiver pair were installed on the same board. The electrical signal from the FPGA transmitter was sent via an interposer to the Firefly transmitter. The interposer was a design feature on previous Serenity versions to allow for easy changes of FPGAs without the need for soldering. The newer versions, not used for the optical tests, will have the FPGA soldered straight onto the board and, thus will suffer less from signal degradation caused by the interposer. In the Firefly transmitter, the electrical signal was converted to an optical signal and sent through the optical fibres. The optical signal then reached the Firefly receiver and was converted back to an electrical signal. The entire setup operated with a 320 MHz clock. An illustration and a picture of the setup are shown in Figure 5.2 and Figure 5.3, respectively.

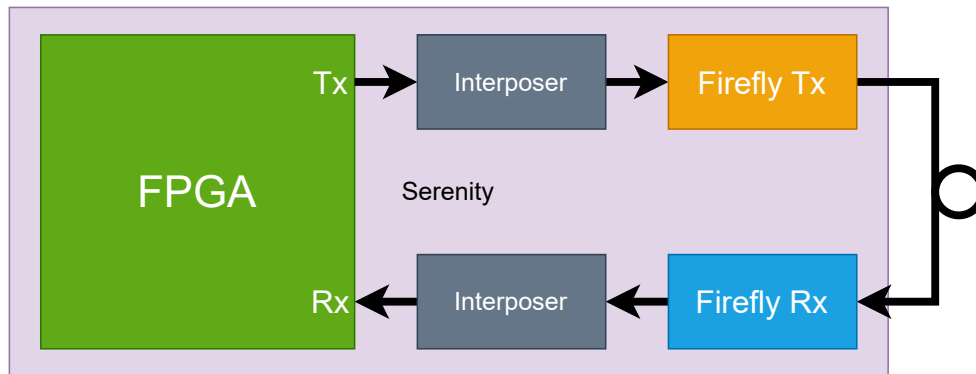


Figure 5.2: The general loopback setup and signal path for the Firefly optical tests.

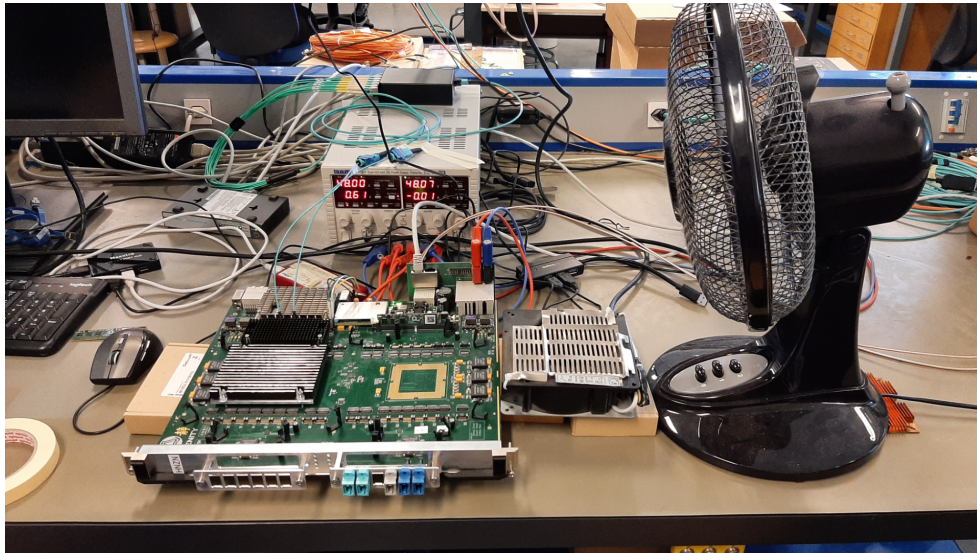


Figure 5.3: The setup of a Serenity board with Firefly transceivers connected in a loopback.

The tests were conducted using Vivado’s Integrated Bit Error Ratio Tester (IBERT), which is specifically designed to evaluate and monitor the high-speed transceivers (GTY) on the FPGA. The test is programmed on the FPGA and includes logic to generate test data, as well as checkers that verify the received data. The IBERT also provides “bathtub” scans, which involve reading out the bits at various points within a clock cycle to determine the range, or open area, where the data can be trusted. Sweeping the bit read-out location over the entire clock cycle, sometimes referred to as a unit interval, and counting the number of errors generates a so-called bathtub curve: a plot of the Bit Error Ratio (BER) as a function of time within the clock cycle. The BER is

defined as

$$\text{BER} = \frac{\text{No. errors} + 1}{\text{No. bits}}, \quad (5.1)$$

where a $\text{BER} < 10^{-12}$ is desired. The BER can never be zero, but it becomes smaller the more bits are measured without any errors. An example bathtub plot is shown in Figure 5.4. The name of the diagram comes from its shape, which tends to have a high BER at the start and end of the clock cycle and a low BER at the middle. This is due to the signal taking time to transition every clock cycle. A good signal has steep edges and a wide opening where the BER is minimal, thus reducing the risk of incorrectly reading out the bit.

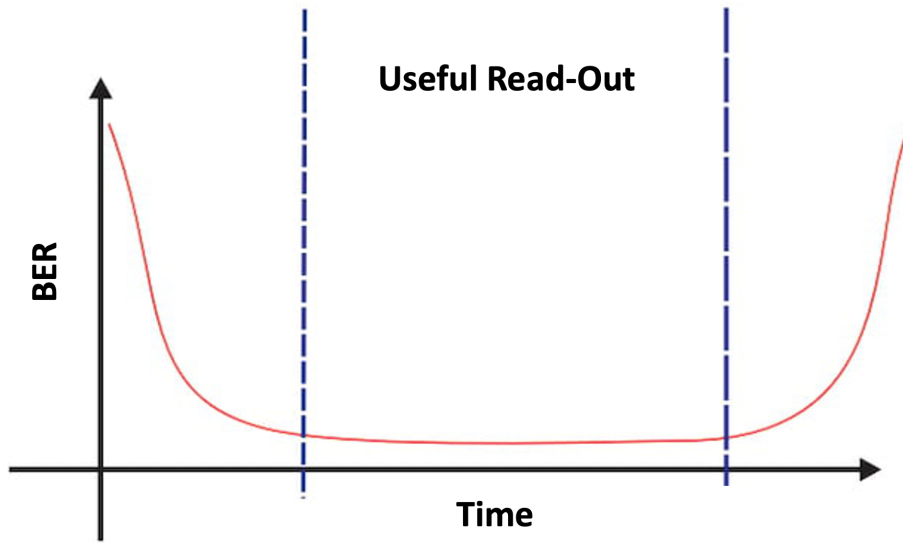


Figure 5.4: An example bathtub curve over one unit interval. Useful read-out occurs within the dashed vertical lines where the BER is at its lowest. The BER is high close to the clock edges.

Thanks to the on-board computer that each Serenity is equipped with, it is possible to remotely program and test the FPGA using Vivado, as well as to change various settings for signal transmission and reception of the GTY transceivers. Remote connections also allow configuration of the Firefly transceivers using the Serenity Management SHell (SMASH) tool [62]. Section 5.4 gives a description of the tests of data integrity stability. These tests were performed by scanning over various FPGA and Firefly operating parameter settings and measuring the performance for each setting combination. Other tests, such as verification of the optical signal from the Firefly transmitter using an optical oscilloscope (Section 5.5) and attenuation scanning to assess signal integrity under attenuation (Section 5.6), are also described.

Pseudorandom Binary Sequence

The BER measurements use the IBERT to transmit and receive a deterministic and periodic bitstream called a Pseudorandom Binary Sequence (PRBS) [63]. This bitstream appears random, making it a suitable test data source. The length of a PRBS k sequence is $2^k - 1$ bits (where $k = 1, 2, 3, \dots$), and it contains every non-zero bit word combination of the length k . The sequence can be repeated until the preferred number of bits has been transmitted. Longer sequences generate more stress on the system under test as they can contain longer series of only 0s or 1s that could negatively affect the CDR performance. As the sequence is deterministic, it will always be the same when repeated, enabling straightforward measurement of bit errors received.

5.4 Parameter Scanning

At high speeds, such as those experienced by Serenity boards in the HL-LHC, the signal is more prone to degradation, making it more difficult for the receiver to interpret the data. Pulses travelling along a fibre will undergo dispersion, which can cause succeeding bits to interfere and blur together. The effect of dispersion increases with high data rates. Interference between bits can also happen when the data contains a long series of only zeros or ones and changes bit value for only a short amount of time before going back to the original value. In this case, the circuit may not have time to reach the full amplitude of the signal and the intermediate bit may not get measured. This is called Inter-Symbol Interference (ISI) [64].

Signals will also be degraded by attenuation, which can be caused by factors such as absorption, scattering, and bending of the optical fibre. Another source of signal degradation is crosstalk between neighbouring channels, meaning that a signal transmitted through one wire may interfere with another channel if they are located closely together.

One way to improve signal quality and mitigate signal degradation is to emphasize and equalize the electrical signals. Emphasis on the signal is performed during transmission to avoid any anticipated noise by boosting the parts of the signal that are most likely to disappear. Equalization is performed when receiving the signal and removes any distortion from the transmission to restore the original signal. Both the FPGA and Firefly provide configurable parameters to manipulate the electrical signal, described further in Sections 5.4.1-5.4.2. In the parameter scanning process, various values of transceiver settings were tested by conducting a BER test and bathtub scan for each setting combination. The scans were performed using Vivado's IBERT (Section 5.3), and each BER scan ran until around 10^{12} bits had been measured, as CMS requires a BER of $< 10^{-12}$ or lower, thanks to protocols for when errors occur. Each bathtub was made using 10^6 bits for each read-out time value for a faster measurement. An acceptable bathtub has an opening that is $> 30\%$

at $\text{BER} < 10^{-12}$, thus an extrapolation of the bathtub curve is made down to 10^{-12} . From the scans, the settings that gave the best results could be determined, as well as how robust the signal is for variations in the settings. The tests were performed separately for the Tx and Rx paths. However, it was noticed that the Rx settings only had a small impact on the performances thus no such measurements were performed for the Beta parts.

5.4.1 FPGA Transceiver Parameters

Transmitter Pre-Cursor and Post-Cursor

Both the pre- and the post-cursor are parameters that emphasise or de-emphasise the signal leaving the FPGA transmitter by changing the part of the signal that is more likely to go undetected, thus avoiding issues of bit transitions being missed. The Tx pre-cursor changes the electrical signal compensating for dispersion, while the post-cursor compensates for previously sent bits to avoid ISI from distortion of the signal amplitude.

Transmitter Differential Swing

In order to reduce the impact of signal noise and distortion, a differential signal consisting of two signals with opposite amplitudes is used in the electrical paths. The final output signal is obtained by taking the difference between these two signals. This approach helps to cancel out much of the noise that would affect both signals similarly. The differential swing refers to the difference between the negative and positive amplitudes of the signal. The amount of differential swing could be changed on the FPGA transmitter going towards the Firefly. A larger differential swing can make the signal more distinguishable, but it also increases power consumption. Moreover, if the value is too large according to the product specifications, it may damage the Firefly and decrease its lifespan.

Receiver Termination Voltage

Receiver termination refers to the resistance or impedance that is connected to the input of a receiver. Matching the impedance of the termination with the characteristic impedance of the wire prevents signal reflections which would interfere with the signal, and ensures maximum power transfer. The FPGA has an adjustable Rx termination resistance which can be modified by changing Rx termination voltage.

5.4.2 Firefly Parameters

Equalization

The Tx module is equipped with equalizers to restore any loss in the electrical signal and signal shape that occurred during transmission from the FPGA to the Firefly. Additionally, it can compensate for any distortion or noise that may have affected the transmitted electrical signal.

Pre-emphasis

Pre-emphasis is a signal processing technique where the signal is intentionally overdriven at the beginning of a transition. The pre-emphasis changes the signal on the Firefly Rx to boost the high-frequency transitions such that they do not get lost once reaching the FPGA receiver.

Amplitude

The amplitude parameter is set on the receiver and specifies what the electrical signal amplitude should be once it exits the Firefly.

5.4.3 Results

The main measurements of the 12-channel Beta parts and the 4+4-channel part are shown here. For additional measurements and results of the Alpha parts, see Appendix B.

The outcome of the parameter scan is displayed as a multi-dimensional histogram, where each bin represents a combination of up to four parameters, and each axis corresponds to one of those parameters. The colour and count in each bin both show the number of links that recorded zero errors during the measurements, i.e. indicating the total number of good links. In the parameter scan plots, there are occasionally also small pink squares inside each bin in a 4×3 (4×2) grid for the 12-channel (4+4-channel) parts, denoting which of the 12 (8) links that showed errors.

Transmitter Parameter Scan

The transmitter parameter scans involved modifying various parameters such as differential swing, pre- and post-cursor on the FPGA, as well as the equalization on the Firefly. In the case of the bidirectional 4+4-channel Firefly transceiver, the receiver's default values were set to a termination voltage of 800 mV, pre-emphasis of 0 dB, and amplitude set to 300 mV. The results of the scan are presented in Figure 5.5. The performance of this Firefly is satisfactory as there are many setting combinations which result in no errors, particularly when the post-cursor is set to 0 dB. For a transceiver to be regarded as functional, at least one setting combination should not show any

errors. However, the more setting options that give good results the better, as it implies a more robust system. The bidirectional Firefly is quite resilient to changes in the other three settings (pre-cursor, differential swing, equalization). A bathtub curve for a well-performing setting of this Firefly can be seen in Figure 5.6. The bathtubs show a decent range in time where the number of errors is minimal, i.e. a large open area.

The default receiver values used for the 12-channel Beta parts were a termination voltage of 800 mV, a pre-emphasis of 2 dB, and an amplitude set to *Medium*. Most of the parameter space showed good signal integrity, as depicted in Figure 5.7, except for the part ICjul22d and ICjul22f which had one link fail using every combination of settings. The investigation into the issue with ICjul22d is described in detail below. However, all bathtub scans for ICjul22d still showed a large open area for all links, even for the failing one, as shown in Figure 5.8. This may be explained by the parameter scan measuring more bits (10^{12} bits compared to 10^6 for each read-out time value for the bathtub scans). After part ICjul22f had been measured, it was later discovered that the part was measured using a broken adapter between the optical fibres coming from the Firefly transmitter and receiver for one of the channels, which caused the poor BER in Figures 5.7-5.8. The broken adapter was later changed for future measurements. Similar parameter and bathtub scan results were obtained for the same transceivers on a different Serenity board, as described in Appendix B.

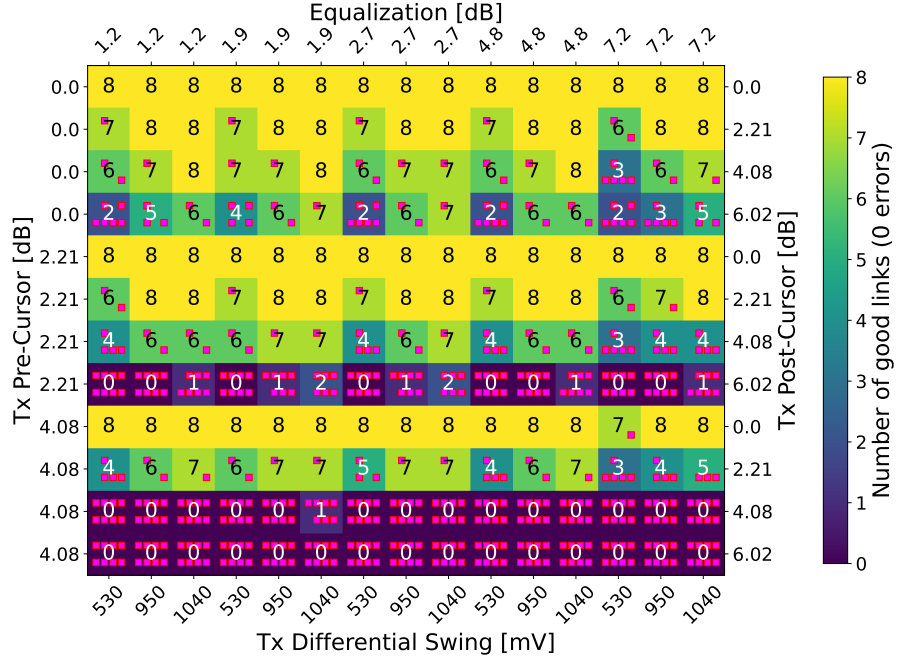


Figure 5.5: Transmitter parameter scan results for the 4+4-channel Firefly. The scan was performed on Serenity 03.

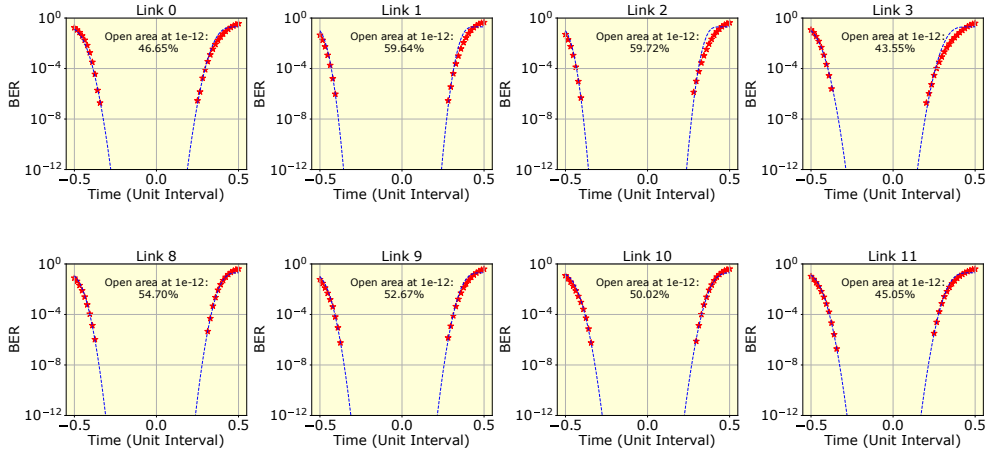
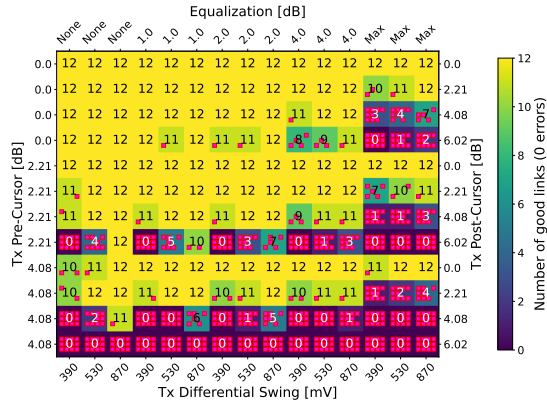
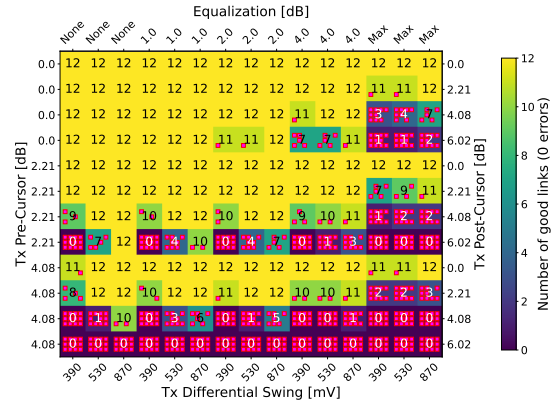


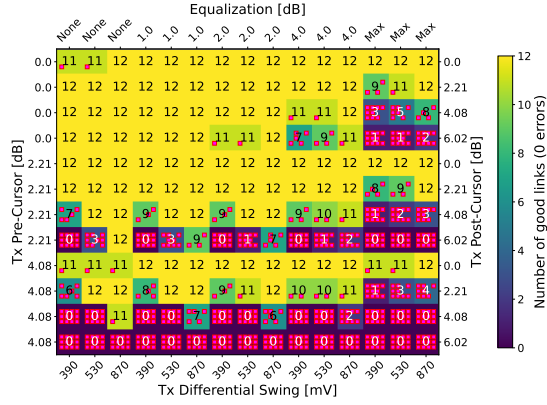
Figure 5.6: Bathtub scans of all links on the 4+4-channel Firefly with the following Tx settings: pre-cursor of 4.08 dB, post-cursor of 0 dB, differential swing of 1040 mV, and equalization of 3.7 dB. The stars correspond to the points of measurement, and the dashed lines are the (complementary) error function fits of the data. An estimation of the open area at a BER value of 10^{-12} for each scan is also given. The measurement was performed on Serenity 03.



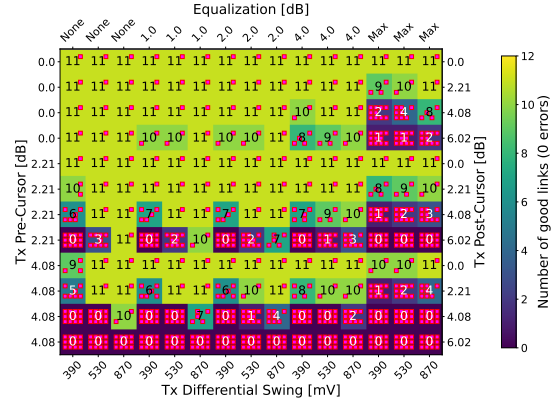
(a) ICjul22a.



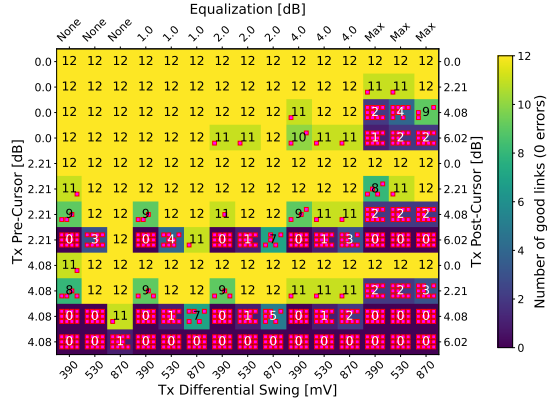
(b) ICjul22b.



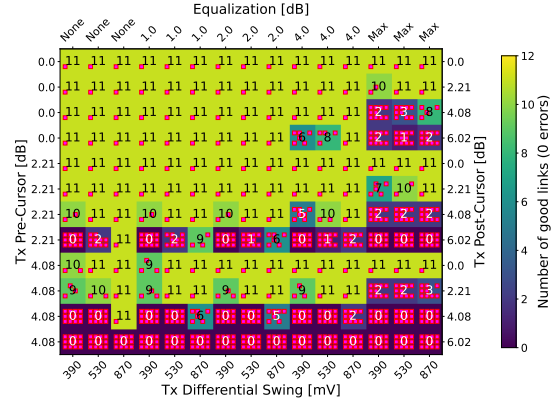
(c) ICjul22c.



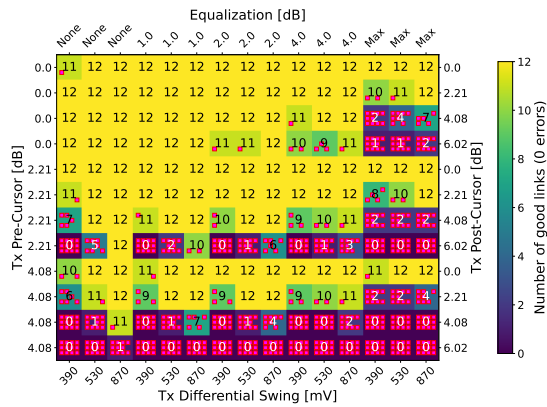
(d) ICjul22d.



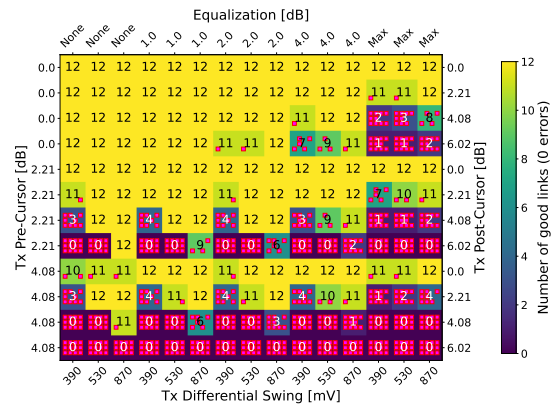
(e) ICjul22e.



(f) ICjul22f.



(g) ICjul22g.



(h) ICjul22h.

Figure 5.7: Parameter scan results of the transmitter settings for the 12-channel Beta parts. The scans were performed on Serenity 04.



Figure 5.8: Bathtub scans for the 12-channel Beta parts with the following well-performing Tx settings: pre-cursor of 4.08 dB, post-cursor of 0 dB, differential swing of 1040 mV, and equalization of 3.7 dB. The stars correspond to the points of measurement, and the dashed lines are the (complementary) error function fits of the data. An estimation of the open area at a BER value of 10^{-12} for each scan is also given. Open areas $< 30\%$ are shown in red. The scans were performed on Serenity 04.

Receiver Parameter Scan

The parameter scans for the receiver involved adjusting the termination voltage on the FPGA, and the pre-emphasis and amplitude on the Firefly. The default settings for the bidirectional 4+4-channel Firefly transceiver were set to the following values: pre-cursor and post-cursor set to 0 dB, differential swing set to 950 mV, and equalization set to 1.3 dB. As shown in Figure 5.9, the receiver parameters were found to be robust against changes, and the majority of the setting combinations resulted in a high number of good links.

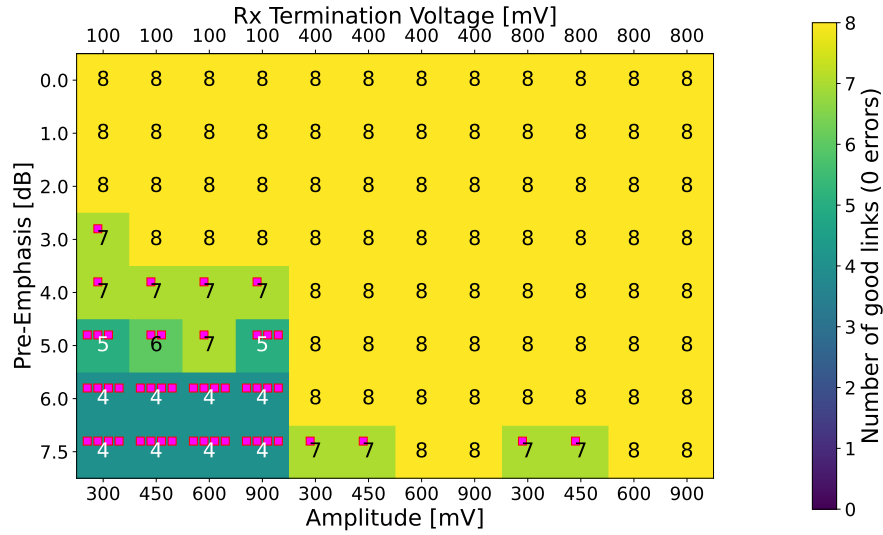


Figure 5.9: Receiver parameter scan results for the 4+4-channel Firefly. The scan was performed on Serenity 03.

Receiver parameter scans were also conducted on 12-channel Alpha parts and showed similar behaviour. Because the Rx parameters had a minor impact on the signal transmission performance, receiver scans were not conducted for the 12-channel Beta parts.

Clock and Data Recovery Issue

Link 3 on part ICjul22d consistently showed errors (Figure 5.7d), regardless of the parameter combination used. Turning off the CDR on the receiver resolved the issue and the link performed well, as seen in Figure 5.10. The same behaviour was observed on a second Serenity board as well. The CDR

is an optional setting to help better match the clock edge with the data edge as noise will be more prominent at 25 Gb/s. Since the CDR is designed to improve performance and stability, this unwanted behaviour is unusual and there is currently no explanation for it. It could be that the CDR is tuned to the wrong frequency. Such problems will be deemed unacceptable for the back-end installation. The issue has been reported to the manufacturer for further investigation, but for the remaining tests, the CDR was kept off. Unfortunately, the CDR cannot be turned on or off for individual links, but the impact on the other links seemed insignificant with the CDR off.

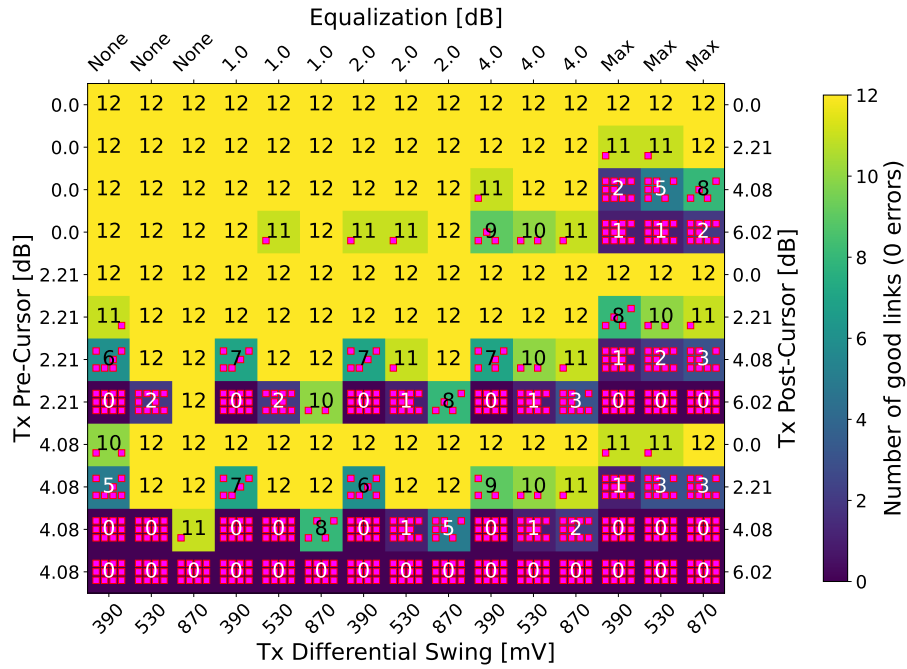


Figure 5.10: Transceiver parameter scan results for ICjul22d with the CDR turned off. The scan was performed on Serenity 04.

Figure 5.11 displays bathtub scans with the CDR turned off, and all the bathtubs remain open, albeit with a slightly smaller opening. Some links are highlighted in red as the opening is now below a threshold set at 30%. However, comparing with the previous measurements with the CDR on, it can be observed that the opening was just above 30% and now just below 30%, thus the difference is minimal.

Similar issues were noted when performing the optical scope and attenuation measurements, which again could be solved by turning the Rx CDR off.

a

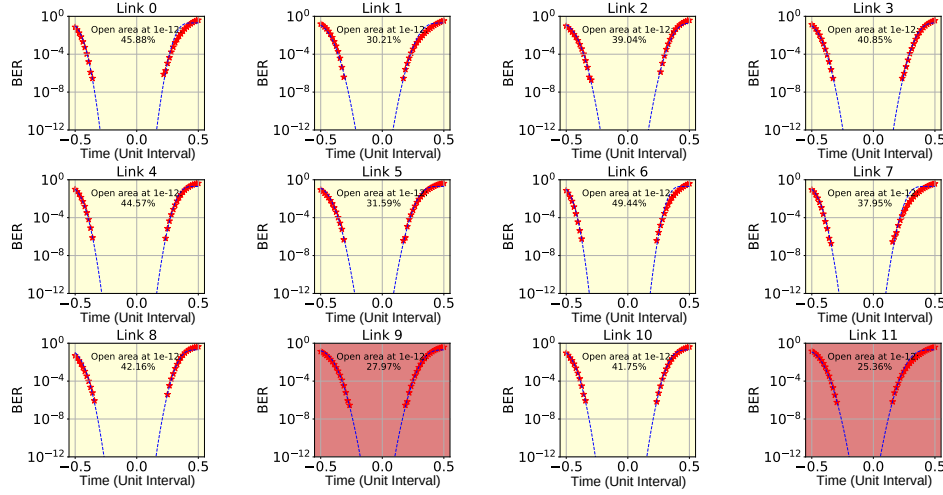


Figure 5.11: Bathtub scan for ICjul22d with the CDR turned off using a well-performing Tx setting: pre-cursor of 4.08 dB, post-cursor of 0 dB, differential swing of 1040 mV, and equalization of 3.7 dB. An estimation of the open area at a BER value of 10^{-12} for each scan is also given. Open areas $< 30\%$ are shown in red. The measurements were performed on Serenity 04.

5.5 Optical Oscilloscope Measurements

To assess the optical signal quality from the Firefly transmitter, an optical oscilloscope (Keysight N1092C) was utilised. Each link was evaluated individually using the bit sequence PRBS7 with an 1.6 GHz oscilloscope input clock as per the oscilloscope's specifications. The data transmission rate remained at 320 MHz. The experimental configuration is depicted in Figure 5.12.

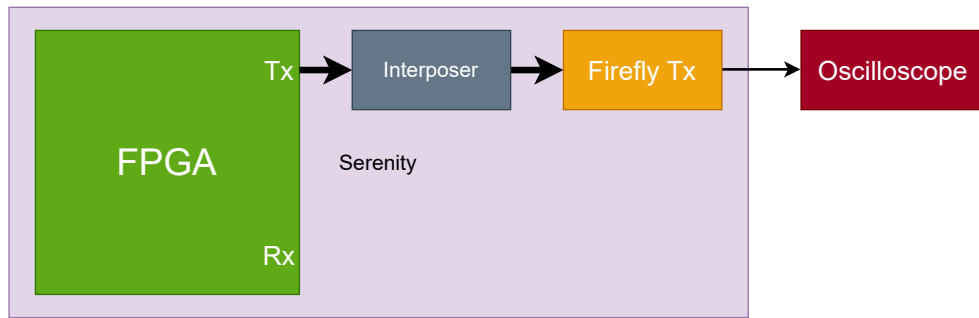


Figure 5.12: The setup and signal path for the optical oscilloscope measurements of the Firefly transmitters. The thin line connecting the oscilloscope symbolises a single channel. The remaining channels are indicated with a thick line.

The optical oscilloscope measurements generate “eye” diagrams, which are composed of numerous signal transitions (going from binary 0 to 1, 1 to 0, and remaining at 0 or 1) sampled on top of one another. When the oscilloscope has accurately locked into the PRBS7 data stream, it will produce an eye diagram that depicts all the various paths the signal can take during a single clock cycle, as shown in Figure 5.13. The eye is the area between the signal and should be completely empty. A larger eye is preferable as it makes it easier to differentiate between binary 0s and 1s. Conversely, a smaller eye can indicate, for example, a noisy signal, ISI, or poor clock synchronisation. The eye diagrams also reveal whether the signal transitions are fast or slow, with the former being more desirable. Eye scans in Vivado can easily be performed on the FPGA receiver side if using the default setup (Figure 5.2). Such scans would provide an overview of the full path. However, using the optical oscilloscope enables performance measurements of the optical transmitters. The following measurements of interest can be obtained from the eye scans, and are evaluated by the oscilloscope:

- Optical Modulation Amplitude (OMA)
- Extinction ratio (r_e)
- Eye width
- Eye height
- Jitter

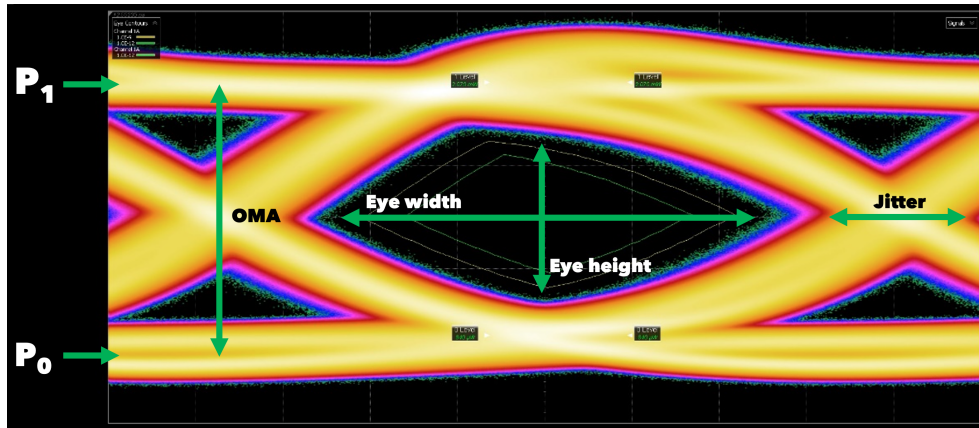


Figure 5.13: An example eye diagram with time on the horizontal axis and power on the vertical axis, along with the definitions of various optical signal characteristics.

Let P_0 and P_1 be defined as the power in milliwatts at the 0- and 1-level, respectively. The OMA_{mW} in milliwatts, as well as the extinction ratio r_e , are

given by

$$\text{OMA}_{\text{mW}} = P_1 - P_0, \quad (5.2)$$

$$r_e = \frac{P_1}{P_0}. \quad (5.3)$$

Equations 5.2-5.3 are the most important values in this study, as they indicate how well separated a 0 bit is from a 1 bit. Acceptable values are given by the manufacturer for each optical transceiver, e.g. for the Firefly beta parts the OMA should be within -3 dBm to 3 dBm and the r_e should be greater than 3 dB. The width of the eye refers to the maximum horizontal opening where no data points are found in units of time. The eye width is related to the open area obtained from bathtub scans, i.e. the region within a unit interval where no errors were measured, thus was not studied further with the oscilloscope. The eye height is the maximal vertical opening where no data points are found in units of power and must be non-zero to be able to separate the 0s from 1s. Both of these quantities need to be non-zero for the eye to be considered open, thus a functioning transmitter will create an eye-shaped area where no data point is located. Additionally, the crossing points of the signal from 0 to 1 and vice versa exhibit a certain amount of horizontal spread, known as jitter. Jitter is defined as the Root Mean Square (RMS) of the displacements from the true signal in time, and can not be too large or the eye will be closed which would yield errors. The larger the eye, the lower the risk of errors. There are no specified requirements of the eye width, eye height, or jitter as the quantities that matter are that the BER is $< 10^{-12}$ and that the bathtub open area at that BER value is $> 30\%$.

In optical communications, power is often expressed in decibel-milliwatts (dBm) rather than milliwatts. The conversion between power in milliwatts and decibel-milliwatts can be done using the following

$$P_{\text{dBm}} = 10 \log_{10} \left(\frac{P_{\text{mW}}}{1 \text{ mW}} \right) \iff P_{\text{mW}} = 1 \text{ mW} \cdot 10^{\frac{P_{\text{dBm}}}{10}}. \quad (5.4)$$

5.5.1 Results

In this section, we will present measurements obtained directly from the optical oscilloscope for the Beta parts, and some measurements from the bidirectional 4+4-channel parts. Initially, the eye diagrams for all links on all parts were satisfactory, and an example of the eye diagrams from ICjul22c is shown in Figure 5.14. However, some parts ceased to function correctly after being permanently damaged during the Firefly testing, which will be described further at the end of this section.

The following default values were used for the oscilloscope measurements of the Beta (4+4-channel) transmitters: pre-cursor of 0 dB (0 dB), post-cursor of 0 dB (0 dB), differential swing of 870 mV (950 mV), and equalisation of 0 dB (1.3 dB). Note that these values do not affect the amplitude of the optical signals.

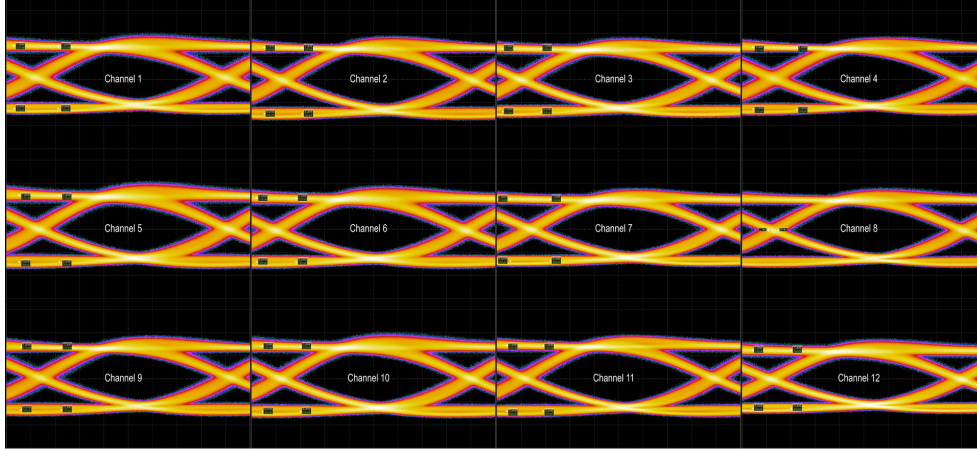


Figure 5.14: Eye diagrams of all links of ICjul22c. Channel number 1 corresponds to link 0 and so forth. Measurements were performed on Serenity 04.

Optical Modulation Amplitude

The Firefly transmitter's OMA is displayed in Figures 5.15-5.16, and they all fall within the component specifications of -3 dBm to 3 dBm. The minimum OMA limit of -3 dBm is the lowest acceptable limit in the worst-case end-of-life scenario.

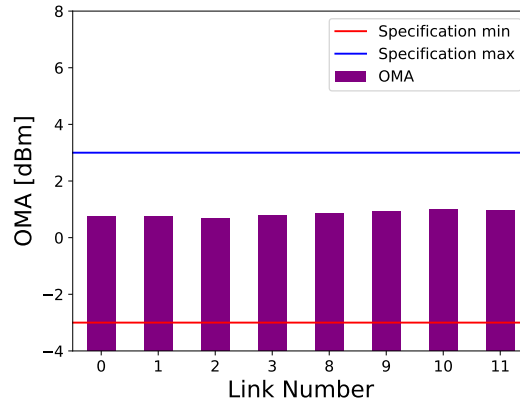
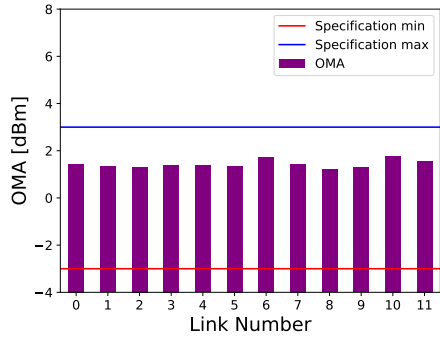


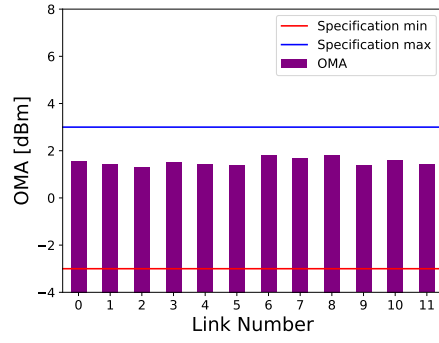
Figure 5.15: The transceiver OMA for all links of the 4+4-channel Firefly. Measurements were performed on Serenity 04.

Extinction Ratio

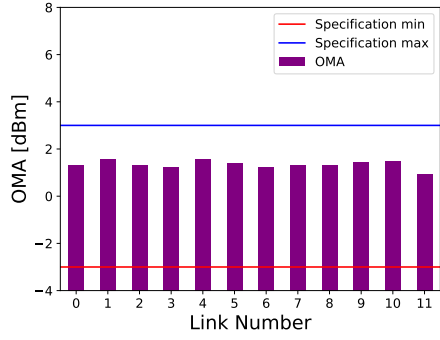
Figure 5.17 shows the measured r_e for each part and link. All parts are above the Firefly minimum extinction ratio specification of -3 dB. However, link 7 on ICjul22g has an exceptionally high extinction ratio compared to the other links. Upon further investigation, it was discovered that the part may have been damaged and would later become unusable, which will be discussed in detail later.



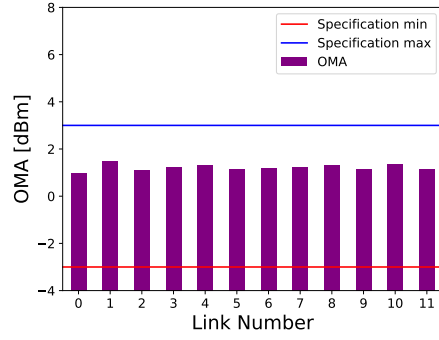
(a) ICjul22a.



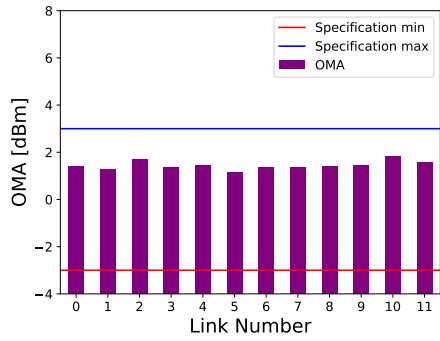
(b) ICjul22b.



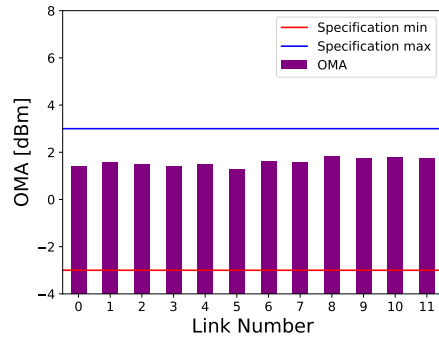
(c) ICjul22c.



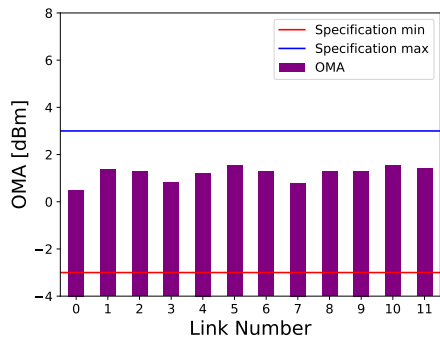
(d) ICjul22d.



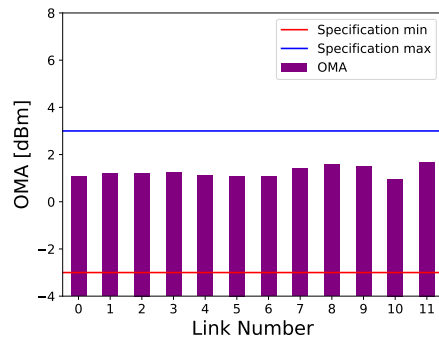
(e) ICjul22e.



(f) ICjul22f.

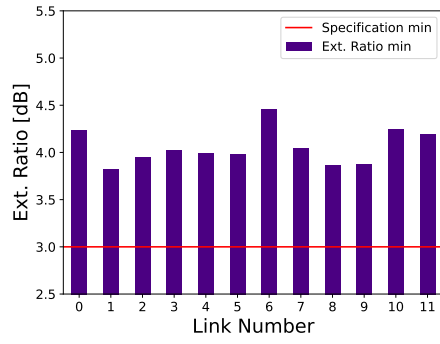


(g) ICjul22g.

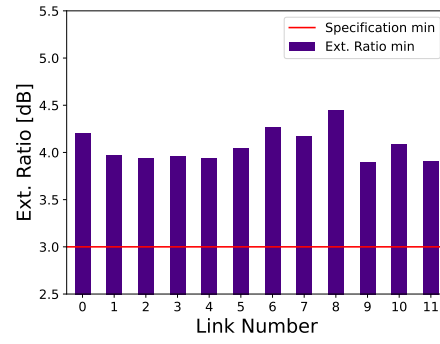


(h) ICjul22h.

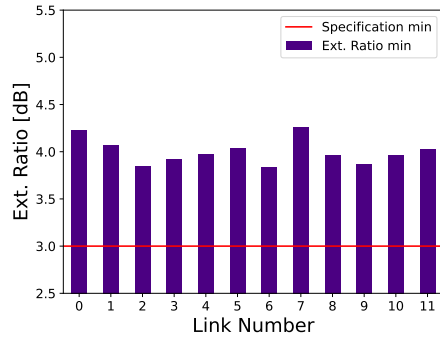
Figure 5.16: The transceiver OMA for each 12-channel Beta part. The blue and the red horizontal lines represent the minimum and maximum OMA according to the specifications, respectively. The measurements were performed on Serenity 04.



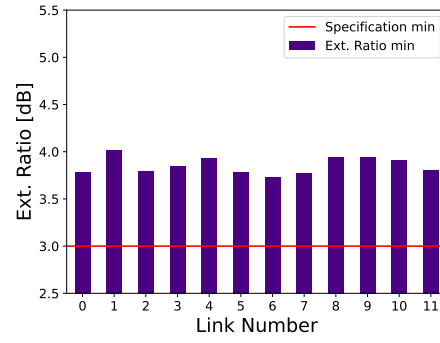
(a) ICjul22a.



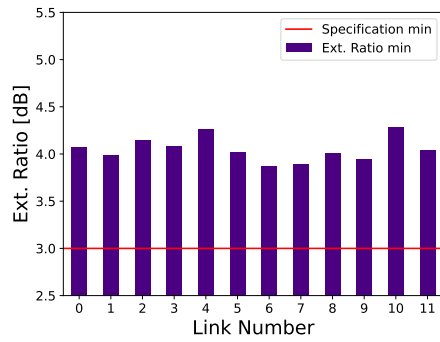
(b) ICjul22b.



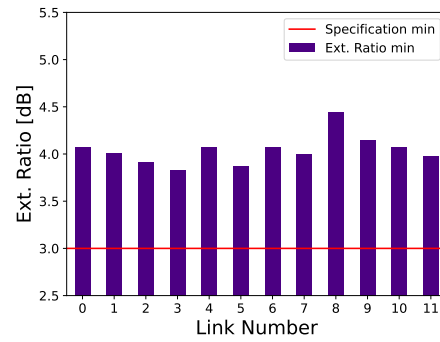
(c) ICjul22c.



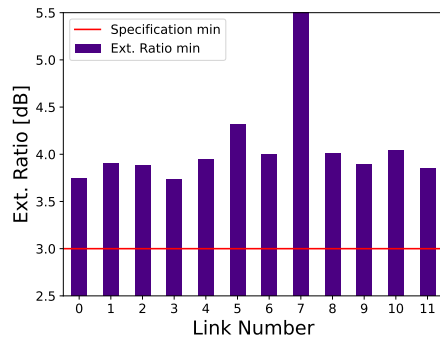
(d) ICjul22d.



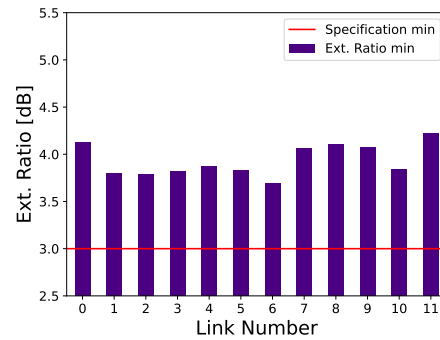
(e) ICjul22e.



(f) ICjul22f.



(g) ICjul22g.

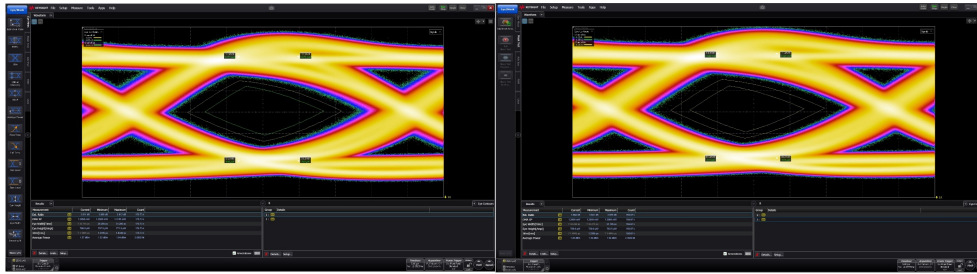


(h) ICjul22h.

Figure 5.17: Extinction ratio for the 12-channel Beta parts. The red horizontal line represents the minimum value according to Firefly specifications. The scans were performed on Serenity 04.

Pseudorandom Binary Sequence Comparison

A comparison between the performance of a single channel (channel 2 on ICJul22e) when using PRBSs of two different lengths, PRBS7 and PRBS31, was also conducted. The eye was measured for over 15 hours in both cases to make sure enough data had been gathered to distinguish any differences. The resulting eyes can be seen in Figure 5.18. There are no noticeable differences in the shape of the eye, and the jitters are comparable: 1.400 ps and 1.419 ps for PRBS7 and PRBS31, respectively. This implies that the Firefly Tx CDR is working properly and is robust against long sequences without bit changes which occur when using PRBS31.



(a) PRBS7, 17 hours.

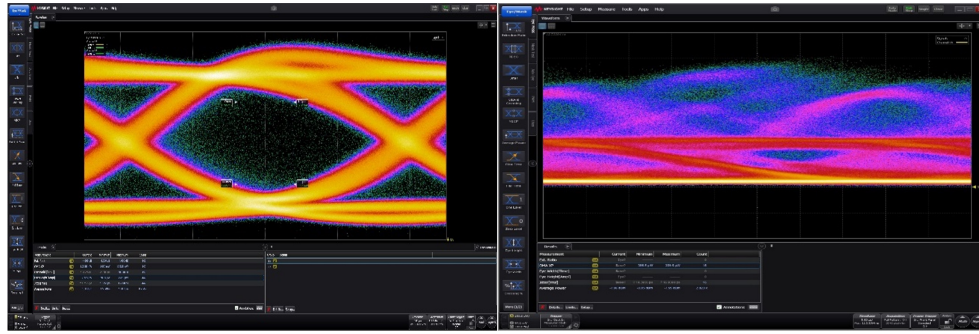
(b) PRBS31, 15 hours.

Figure 5.18: Eye diagrams from the optical oscilloscope of channel 2 on ICJul22e when transmitting PRBS7 and PRBS31 data for more than 15 hours.

Dead Channels

After multiple tests had been performed, it was discovered that certain channels on previously functional parts were no longer operating correctly as high values of BER were measured. During remeasurements with the oscilloscope, it became apparent that the issue was caused by the Firefly transmitters. Eye diagrams that had previously demonstrated acceptable eye shapes now indicated poor performance, which suggested that these channels were damaged or prematurely worn out. Figure 5.19 shows the eye diagrams for these affected channels, revealing that the Firefly transmitter still transmitted light, however, the signal was no longer usable as the eye was no longer open. Link 0 on ICjul22e still displayed the shape of an eye, but there was significant noise within the eye which would produce a significant amount of errors when transmitting data. No hits measured in the centre of the eye are allowed as the eye will be deemed closed. Link 7 on ICjul22g did not show any resemblance to an eye. Both damaged links did not produce any meaningful results for the rest of the tests.

Two theories were initially proposed by Samtec to explain the catastrophic failure of the Firefly transceivers: insufficient anti-static procedures during



(a) ICjul22e, link 0.

(b) ICjul22g, link 7.

Figure 5.19: Eye diagrams from the optical oscilloscope of two channels that no longer work. Measurements were performed on Serenity 04.

installation, or incorrect power on/off procedures. Given that no other Alpha part had previously been damaged in this manner, these explanations seemed unlikely. Nevertheless, to prevent any potential issues, anti-static wristbands were worn during future handling of the parts, and the voltage was ramped up/down when powering the Firefly transceivers on/off as per Samtec’s revised instructions. Later, it was discovered by the manufacturer that the dead channels might have been caused by problems with the VCSELs during the fabrication process. The VCSEL supplier has updated their manufacturing process to address this issue, and both Samtec and the VCSEL supplier are currently conducting rigorous testing on the VCSELs.

5.6 Attenuation Scanning

An essential characteristic of any communication system is its ability to tolerate attenuation. As signals traverse through cables and optical fibres, they experience attenuation which increases as the distance travelled increases. Scanning the transceivers at different levels of attenuation provides valuable information on how much signal loss a channel can withstand before it becomes error-prone.

The measurements were carried out using a setup similar to the one described in Section 5.3, except that a variable attenuator (Noyes VOA5) was installed between the Firefly transmitter and receiver, as shown in Figure 5.20. A picture of the setup can be seen in Figure 5.21. The attenuator allowed for control of the attenuation levels in the system, but only one fibre could be attenuated at a time. Each fibre was measured individually, and all parts were measured with all fibres connected to the receiver to account for the maximum amount of cross-talk between the channels. Some parts were also measured with only one fibre connected to the receiver, which avoided the effects of cross-talk in the receiver. As with the previous measurements, a

PRBS31 signal was used, and the measurements continued until either errors were detected or 10^{12} bits had been received.

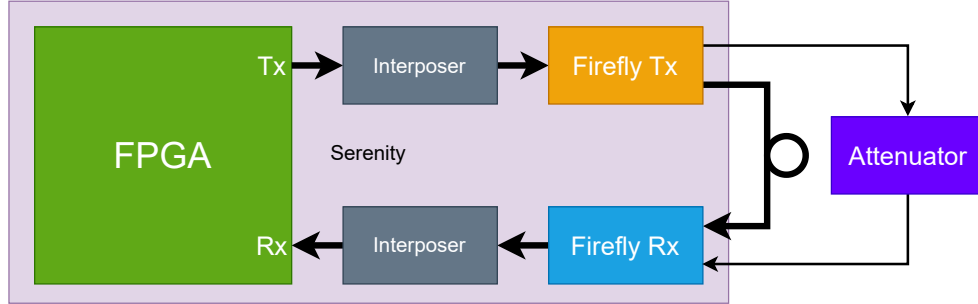


Figure 5.20: The setup and signal path for the attenuation measurements. The thin arrows connecting the attenuator symbolise a single channel. The remaining channels are indicated with a thick line.



Figure 5.21: The setup of a Serenity board with Firefly transceivers connected in a loopback via an attenuator. The picture also shows the optical oscilloscope and the external power source.

5.6.1 Results

In this section, the results of the attenuation measurements conducted on the Beta parts and the bidirectional Firefly Transceivers will be presented. For measurements on Alpha parts, see Appendix A.

The following default values were used for the attenuation measurements of the 12-channel Beta (4+4-channel bidirectional) transceivers: pre-cursor of 0 dB (0 dB), post-cursor of 0 dB (0 dB), differential swing of 870 mV (950 mV),

equalisation of 0 dB (1.3 dB), termination voltage of 800 mV (800 mV), amplitude of *Medium* (300 mV), pre-emphasis of 2 dB (0 dB).

The attenuation scans measured the BER at different levels of attenuation. A higher tolerance for attenuation indicates a more robust system. The figures presented in this section (Figures 5.22-5.23) do not include any attenuation offset caused by the attenuator itself, so the values on the x -axis represent the attenuator setting rather than the actual attenuation. The attenuator gave an additional attenuation of around 1 dB on top of the set value, discussed in Section 5.7. As desired, a $\text{BER} < 10^{-12}$ was achieved at low attenuation. Most parts began showing errors at 5-10 dB of attenuation. Typically, the attenuation of light in an optical fibre is $< 1 \text{ dB/km}$ [59]. This would correspond to a distance much further than any distances in the CMS back-end, which is of the order of 10 m.

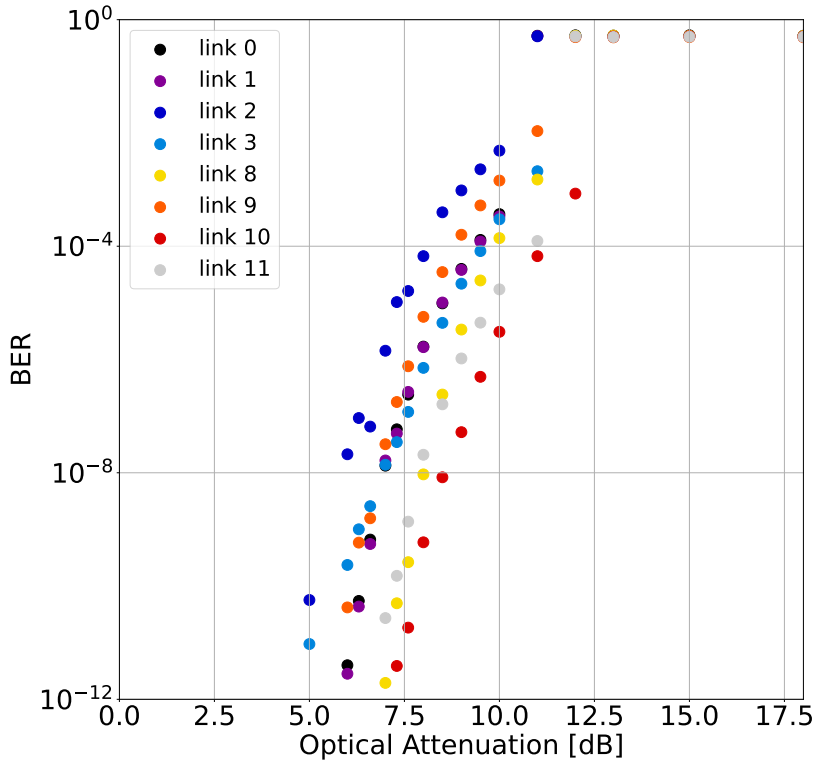
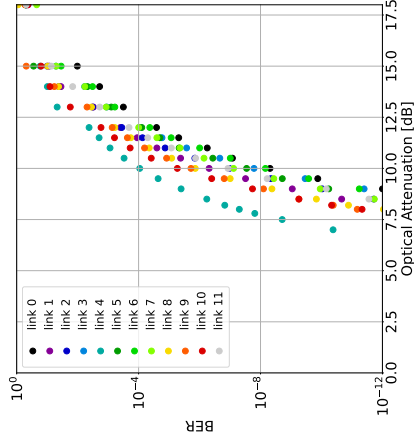
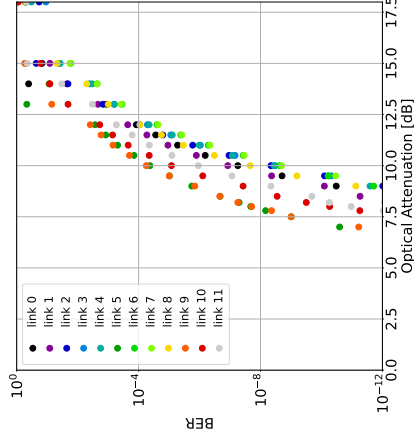


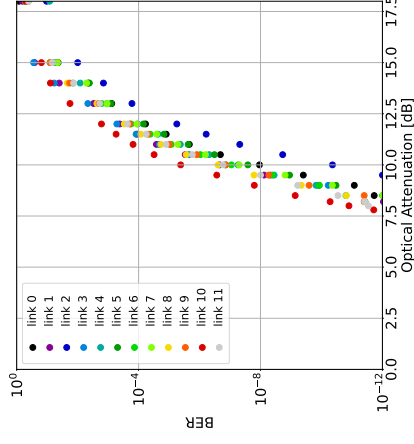
Figure 5.22: The BER as a function of applied attenuation on the 4+4-channel Firefly. Note that the attenuator still gives some attenuation when set to 0 dB which has not been included in this plot. The scans were performed on Serenity 04.



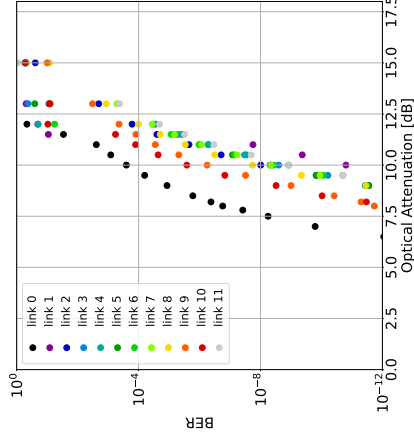
(a) ICjul22a.



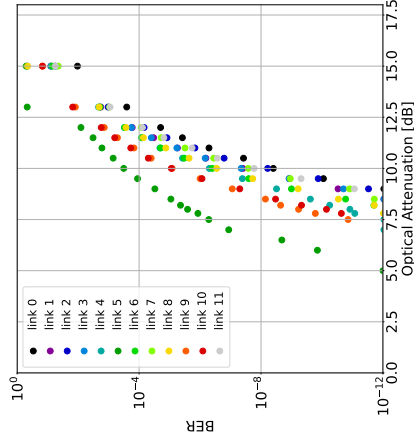
(b) ICjul22b.



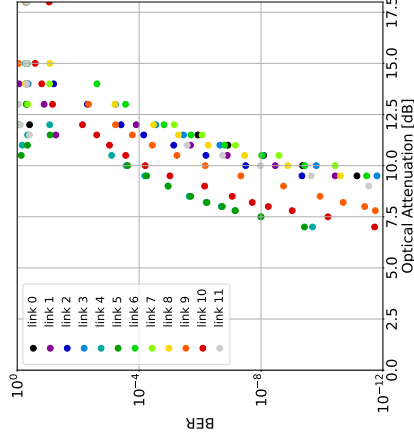
(c) ICjul22c.



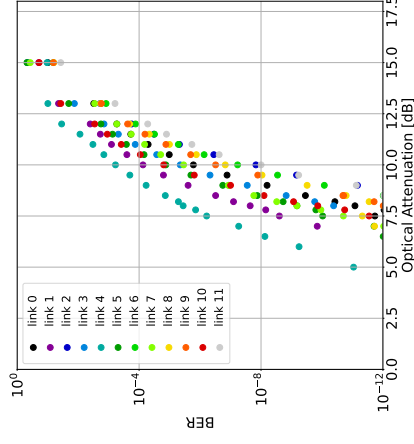
(d) ICjul22d.



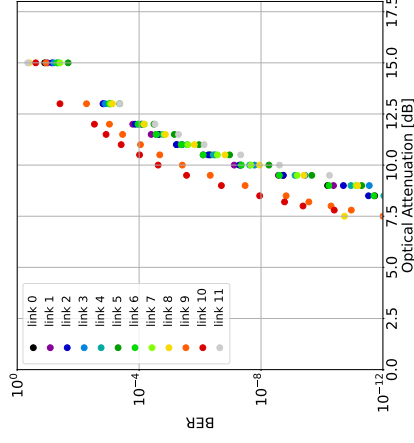
(e) ICjul22e.



(f) ICjul22f.



(g) ICjul22g.



(h) ICjul22h.

Figure 5.23: The BER as a function of applied attenuation on the 12-channel Firefly transceivers. The ICjul22d has CDR turned off. Note that the attenuator still gives some attenuation when set to 0dB which has not been included in this plot. The scans were performed on Serenity 04.

Cross-talk measurements

To investigate the impact of receiver cross-talk on the Firefly performance, measurements were also conducted with only the link being measured connected to the receiver. All other links were still transmitting data. This is in contrast to the default attenuation measurements where all links were connected to the receiver to maximize cross-talk, thus putting more stress on the system. The comparison between these two approaches for the bidirectional 4+4-channel Firefly can be observed in Figure 5.24. As anticipated, the performance was better when not subjected to cross-talk, as the system was able to handle more attenuation before errors were detected.

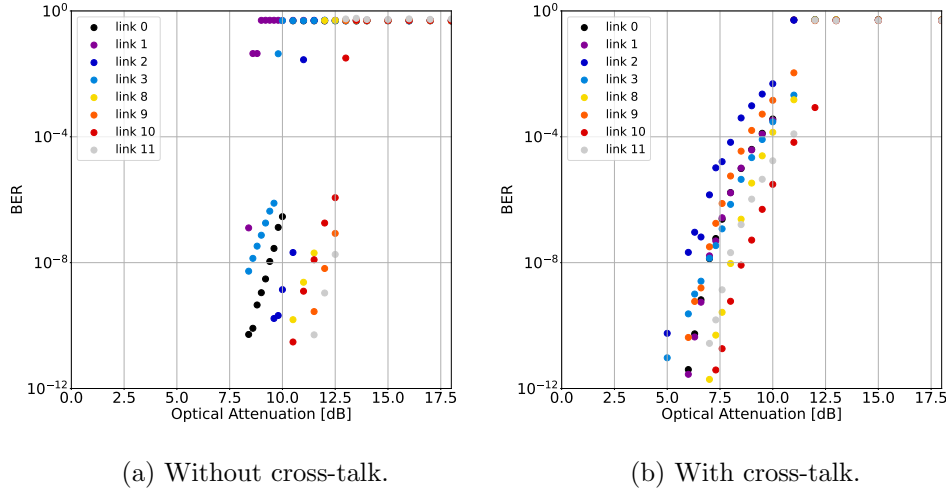


Figure 5.24: A comparison of how cross-talk in the receiver affects the attenuation measurements. The default scans included cross-talk. Measurements were performed on Serenity 03.

Another set of cross-talk measurements on the receiver was performed to identify if the receiver cross-talk occurs in the electrical path or the optical path. As the optical links outside the Firefly tend to be well-separated as individual cables, any cross-talk in the optical path is believed to occur inside or at the interface of the optical transceiver, before equalization, etc. To test this, another attenuation scan was made, but this time all links that were not measured (but still connected) were attenuated to a fixed Tx OMA of -3 dBm between the Firefly Tx and Rx. The attenuation of the links was implemented by adding fixed attenuators until the desired Tx OMA was measured on the optical oscilloscope. Most links originally had an Tx OMA of approximately 1 dBm (Figure 5.16), thus experienced a 4 dB attenuation. Once the correct fixed attenuators had been found for each link, an attenuation scan was performed as usual. If the cross-talk occurred on the electrical path, attenuating

the links in the optical path would not affect any of the results when measuring the BER as a function of attenuation. If it occurred on the optical side, there would be improvements measured corresponding to the amount that was attenuated, i.e. a 4 dB increase of attenuation before errors occurred.

These cross-talk measurements were done on only one spare part, the ICjul22j, and the original extinction ratios and Tx OMA values can be seen in Figure 5.25. An improvement of approximately 2 dB per link was observed for the attenuation measurements in Figure 5.26. This means that cross-talk occurs in the optical path, but cross-talk in the electrical path can not be excluded and would need further investigation. As can also be noticed in Figure 5.26, one link (link 8) always experienced issues. The cause of this was most likely also from a faulty VCSEL, similar to parts ICjul22e and ICjul22g.

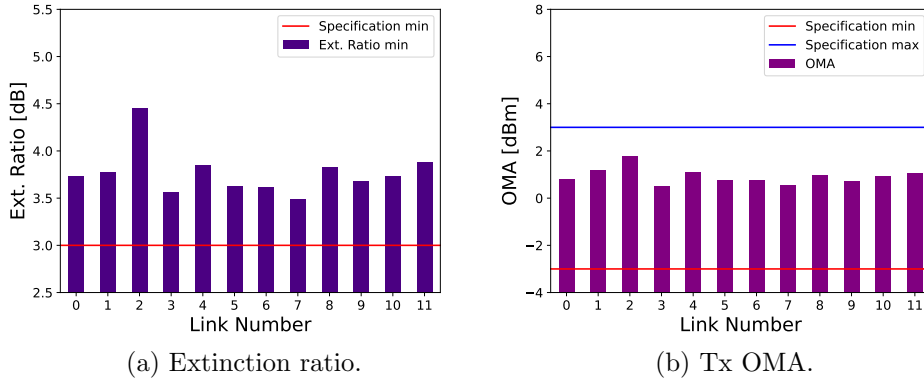


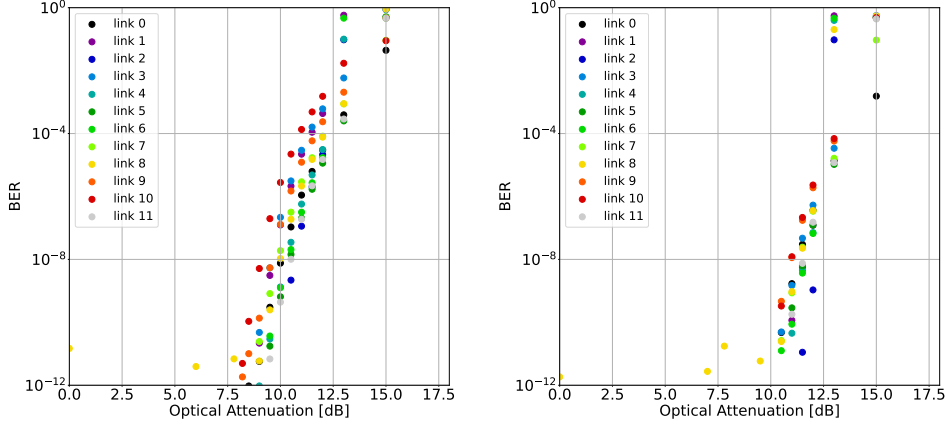
Figure 5.25: The extinction ratio (left) and Tx OMA (right) for the 12-channel Beta part ICjul22j. The blue and the red horizontal lines represent the minimum and maximum values according to the specifications, respectively. Measurements were performed on Serenity 04.

5.7 Receiver Measurements

By combining the optical oscilloscope measurements with the attenuation measurements, it is possible to determine the OMA and average power at the receiver in relation to the amount of attenuation. Since receivers have minimum and maximum specifications for various quantities, it is essential to verify if the receiver performance is within the prescribed range. Since the oscilloscope values are measured at the transceiver, the Rx OMA can be calculated as

$$\text{OMA}_{\text{dBm}}^{\text{Rx}} = \text{OMA}_{\text{dBm}}^{\text{Tx}} - \text{attenuation} - \text{attenuator offset}, \quad (5.5)$$

where the attenuator offset is the attenuation measured when the attenuator was set to 0 dB. To determine this offset, the attenuator set to 0 dB was



(a) Default attenuation measurement. (b) Measurement with fixed attenuators.

Figure 5.26: A comparison of how cross-talk in the receiver affects the attenuation measurements with fixed attenuators (Tx OMA -3 dBm) on the links when not being measured. Measurements were performed using ICjul22j on Serenity 04.

connected between the Firefly transmitter and the optical oscilloscope. The offset was measured as 1.28 dB using a third-generation Alpha part (Alpha3), and this offset was originally used for all measurements. However, it was later discovered that the attenuator did not attenuate each part equally, and each link within a part was attenuated differently, requiring a different offset for each link. The varying attenuation could be due to differences in connections after plugging the fibres in/out and cable bend. The original 1.28 dB offset was used for the 4+4-channel Firefly measurements, but a different method was used for Beta parts, as the Received Signal Strength Indicator (RSSI) values were measured and saved. The RSSI is a quantity, measured by the Firefly itself, of the average power received. By taking the difference between the RSSI values between measurements without the attenuator and with the attenuator at 0 dB, an offset was calculated on a per-link basis. See Section 5.7.1 for a consistency measurement of the RSSI versus the average power measured by the oscilloscope.

To calculate the Rx average power the following relation between the OMA (Equation 5.2) and the extinction ratio (Equation 5.3) was used

$$\bar{P}_{mW} = \frac{1}{2} \cdot \text{OMA}_{mW} \cdot \frac{r_e + 1}{r_e - 1} \implies \bar{P}_{mW}^{\text{Rx}} = \frac{1}{2} \cdot \text{OMA}_{mW}^{\text{Rx}} \cdot \frac{r_e + 1}{r_e - 1}. \quad (5.6)$$

Note that the extinction ratio will stay constant even if the signal is attenuated [65].

5.7.1 Results

Bit Error Ratio vs Receiver Optical Modulation Amplitude

While the attenuation measurements in Section 5.6.1 provide an idea of the maximum attenuation the system can withstand, the Rx OMA offers a better understanding of the receiver's sensitivity and its ability to detect weak signals without errors. A more sensitive receiver is better suited to handle attenuation in the system. The results for the 4+4-channel and 12-channel Firefly receivers are displayed in Figures 5.27-5.28, where most links do not show errors until the Rx OMA drops below -6 dB, thus agree with the product specification. For the bidirectional 4+4 Firefly, two links show errors slightly above -6 dB, but this could be due to an inaccurate estimation of the attenuation offset.

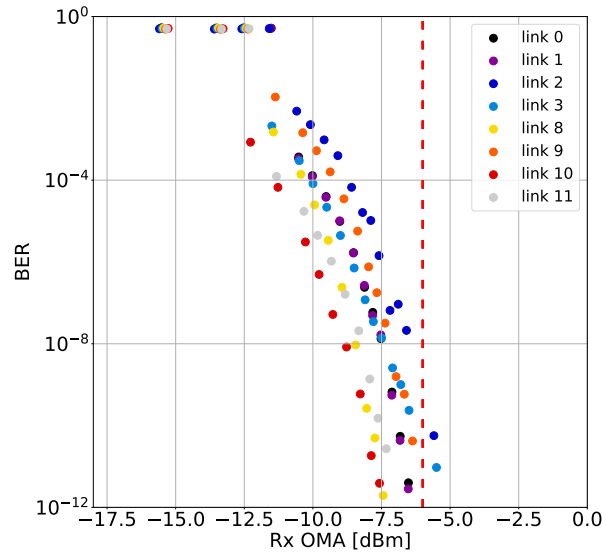
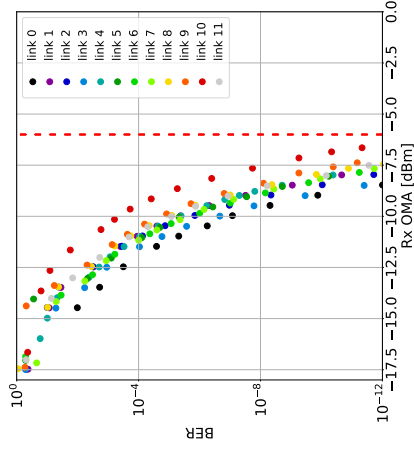


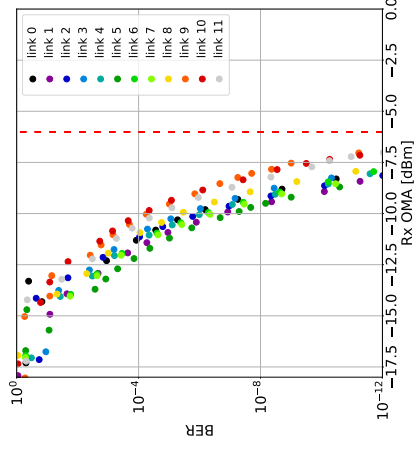
Figure 5.27: The BER as a function of the calculated receiver OMA for the 4+4-channel Firefly. The red vertical line is the minimum Rx OMA value the parts should be able to handle before showing errors according to the specifications. A 1.28 dB attenuator offset was used. The scans were performed on Serenity 04.

RSSI vs average power

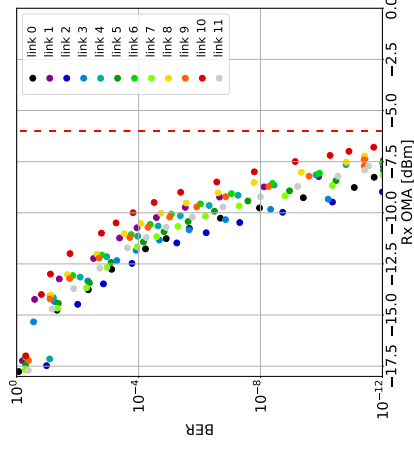
The Received Signal Strength Indicator (RSSI) and average Rx power both measure the same thing, but the former is measured by the Firefly itself while the latter is calculated using previous oscilloscope and attenuation measurements. Thus, a linear one-to-one relationship is expected when plotting the



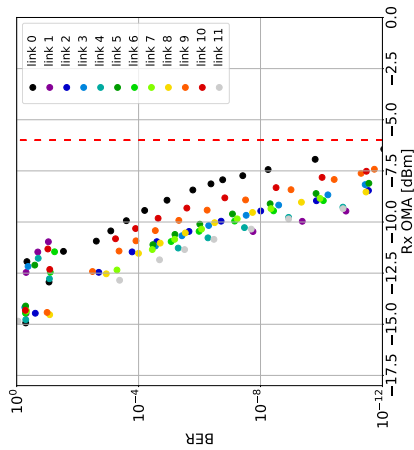
(a) ICjul22a.



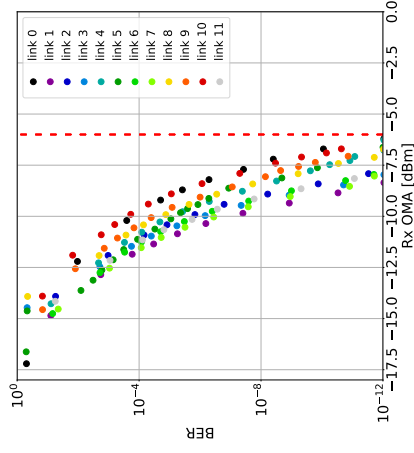
(b) ICjul22b.



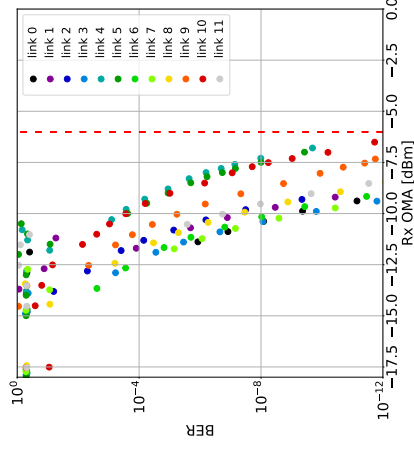
(c) ICjul22c.



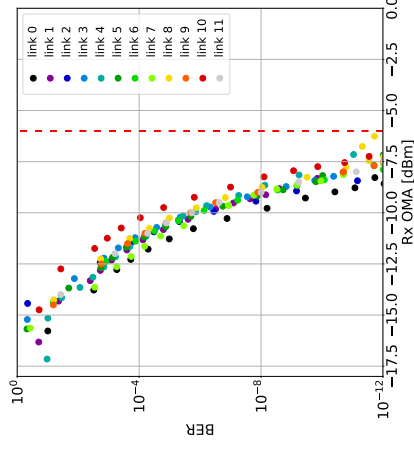
(d) ICjul22d.



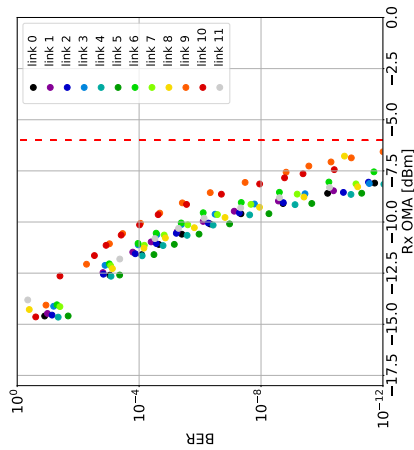
(e) ICjul22e.



(f) ICjul22f.



(g) ICjul22g.



(h) ICjul22h.

Figure 5.28: The BER as a function of the calculated receiver OMA for each 12-channel Beta part. The red vertical line is the minimum Rx OMA value the parts should be able to handle before showing errors according to the specifications. The attenuation offset was calculated using the Firefly RSSI values. The scans were performed on Serenity 04.

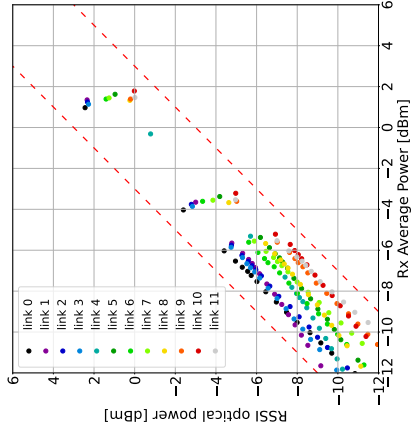
RSSI against the average power for all attenuation values. The RSSI measurements were only taken for the Beta parts. The Firefly transceivers measure the average power/RSSI less accurately compared to the optical oscilloscope, hence a ± 3 dB uncertainty of the RSSI value had been specified for the Beta parts by the manufacturer. Figure 5.29 shows a linear relationship for most parts, with the majority of measurements falling within the ± 3 dB uncertainty. The optical oscilloscope measurements are highly precise, resulting in a small spread in the horizontal direction for the links. However, the RSSI measurements from the Firefly receiver are not as precise, causing the data points to spread out more in the vertical direction.

5.8 Conclusions

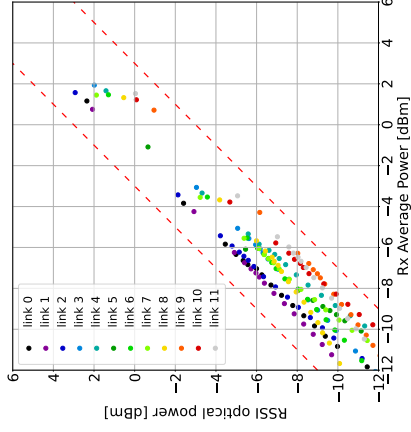
For the CMS detector, high-speed optical transceivers are necessary for data transmission at the HL-LHC. Specifically, the Serenity boards for the Outer Tracker DTC system require 25 Gb/s transceivers to handle the amount of data. To ensure that the transceivers are suitable, performance tests were conducted using eight 12-channel prototype Firefly transceivers and one mature 4+4-channel bidirectional Firefly transceiver, all manufactured by Samtec. A performance scan was made, measuring various combinations of the configuration parameter settings, and most transceivers showed stable performance. However, one link experienced issues with its CDR, which were remedied by turning the CDR off.

Optical oscilloscope measurements were also carried out to ensure that all parts were within product specifications. One link showed odd behaviour during the oscilloscope measurements, and it later completely failed. Another link started getting noisy after multiple tests. Attenuation measurements were also performed, and the majority of the parts satisfied the attenuation requirements.

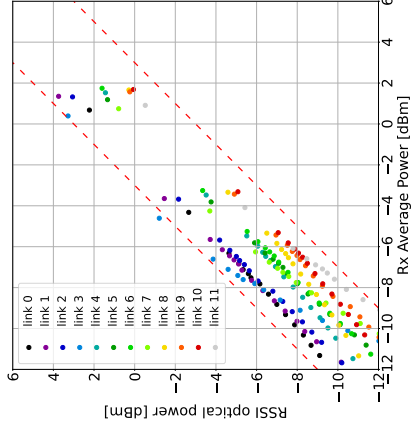
Overall, the tests demonstrated acceptable behaviour from the transceivers. Although three 12-channel prototype parts exhibited issues (excluding the spare part), the vast majority of the links (93 out of 96) worked well. For the final installation of the transceivers in the CMS back-end system, all parts are expected to be fully functional. Communication with the manufacturer is in progress to resolve the issues that were observed with the three faulty channels. These results indicate that the tested transceivers can be used for data transmission at the HL-LHC with reasonable confidence. However, continued monitoring and testing of the transceivers will be necessary to ensure that the observed issues have been addressed and to ensure their reliable performance in the long term.



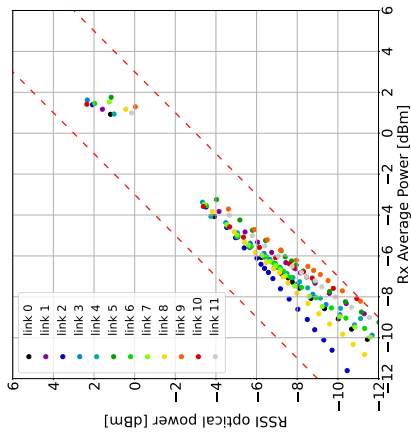
(a) ICjul22a.



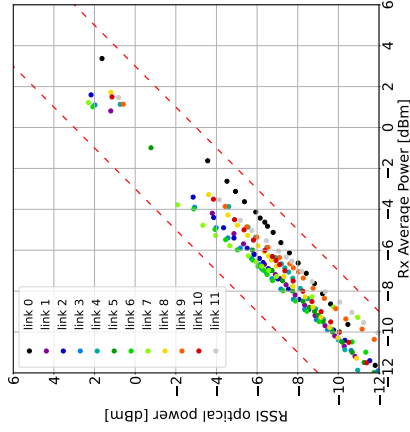
(b) ICjul22b.



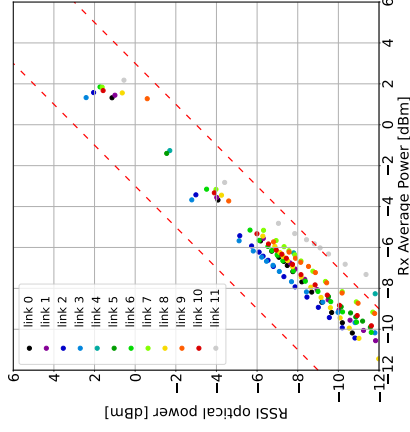
(c) ICjul22c.



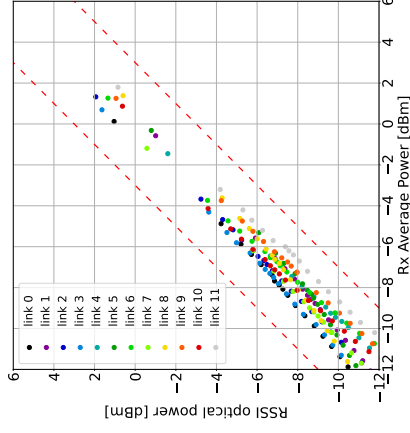
(d) ICjul22d.



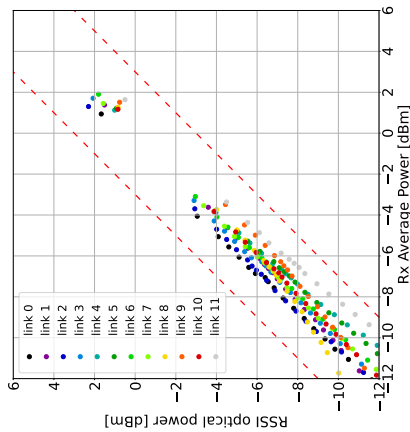
(e) ICjul22e.



(f) ICjul22f.



(g) ICjul22g.



(h) ICjul22h.

Figure 5.29: The RSSI as a function of the calculated average power at the receiver for each 12-channel Beta part. The values should be within the red lines according to specifications. The scans were performed on Serenity 04.

Chapter 6

Search for Light Higgs Bosons from SUSY Decay Cascades

6.1 Introduction

While there has not yet been evidence of BSM physics at the LHC, plenty of the LHC physics potential remains to exploit. Alongside new capabilities from the upgraded detector and the increased integrated luminosity of future runs, innovative analysis techniques are crucial to the discovery prospects of the LHC. These techniques include advancements in identifying the flavour of jets and reconstructing the masses of the particles causing those jets, all of which are challenging tasks due to the intricate nature of jets. Many physics analyses, including searches for BSM physics, include jets in the signal and hence would benefit from these advancements.

Many BSM theories require additional Higgs bosons beyond the already-discovered SM Higgs (H_{SM}), such as SUSY and other two-Higgs-doublet models [66]. The extra particles that come with e.g. SUSY models can then solve issues with excessive fine-tuning of the Higgs mass in the SM. Light neutral Higgs bosons, maybe even lighter than the H_{SM} , could be produced at the LHC and have gone undetected due to small direct production cross sections. As discussed in Section 1.3, much of the Minimal Supersymmetric Standard Model (MSSM) parameter space has already been probed and excluded at the LHC which has set constraints on the properties of any potential SUSY particles. Such constraints can be evaded in the Next-to-Minimal Supersymmetric Standard Model (NMSSM) (also discussed in Section 1.3) when there is a singlino-like Lightest SUSY Particle (LSP) and thus no direct squark to LSP decay, leading to decay cascades. In such scenarios, signal events can have very different properties than typically expected in the MSSM, such as having very little $p_{\text{T}}^{\text{miss}}$ [10].

In this chapter, an NMSSM search for pairs of light Higgs bosons (H_1) produced in supersymmetric decay cascades is discussed. The chapter builds

upon a previously published analysis [12] conducted using CMS Run 2 data. The published analysis focused on the all-hadronic final state scenario where the pairs of light Higgs bosons would decay into boosted pairs of b -quarks which were reconstructed using the anti- k_T (AK) algorithm, i.e. the standard jet clustering algorithm at CMS (Section 2.3.7). The boosted $b\bar{b}$ pairs were reconstructed as wide AK8 jets, such that the $H \rightarrow b\bar{b}$ decay could be represented by a single jet. AK8 jets have twice the radius than the standard CMS AK4 jets, and are also known as fat jets. The properties of the jets were measured in the published analysis using substructure techniques to determine e.g. the mass using the Softdrop algorithm [67], and to identify the $b\bar{b}$ jets using a $H \rightarrow b\bar{b}$ specific jet flavour tagging algorithm called btagHbb¹ [68]. A model parameter space was chosen such that the p_T^{miss} would be significantly suppressed and the boosted Higgs boson would be the main detector signature of the decay. However, no evidence of an excess beyond what is predicted by the SM was found for the chosen benchmark model.

Using the Run 3 data, it is possible to conduct an enhanced analysis of the same signal model using novel techniques for probing the substructure of AK8 jets, such as ParticleNet (Section 2.3.7), which can be used to perform both identification of $b\bar{b}$ jets and mass regression. This chapter explores the possibility of using ParticleNet to improve the published analysis. A comparative study is conducted between the performance of ParticleNet [69] and the previously used mass reconstruction and jet flavour tagging algorithms. The differences in both the background and signal distributions will be presented along with an assessment of the feasibility of using ParticleNet for this analysis. Potential modifications necessary to accommodate the use of the ParticleNet quantities as well as further analysis improvements in preparation for Run 3 will also be discussed.

6.2 Double- b -tagging

A double- b -tagger is an algorithm which identifies AK8 jets originating from a $b\bar{b}$ pair. A double- b -tagger generally returns a discriminator value that indicates whether the jet is likely to contain a $b\bar{b}$ pair. Advancements in machine learning have made implementations of jet flavour tagging algorithms more powerful because of the complex and subtle differences between different jet flavours. Machine learning enables the identification of these differences and improves the accuracy of flavour classification. Some taggers have been trained to mainly classify jets from specific decays, such as the btagHbb, a double- b -tag discriminator for boosted $H \rightarrow b\bar{b}$ decays [35]. Other tagging algorithms are more general and can be used to classify jets from a $b\bar{b}$ pair originating from the decay of an unspecified particle. One example of such a classifier

¹btagHbb is also known as pfBoostedDoubleSecondaryVertexAK8BJetTags

is the ParticleNet double- b -taggers. Overviews of the two double- b -tagger are given below.

6.2.1 btagHbb

The btagHbb double- b -tagger [68] has been used for a couple of years, which at the time it was created showed superior performance over earlier double- b -tagging techniques. It uses a Boosted Decision Tree (BDT), a machine learning algorithm used for classifying data. A BDT takes a set of input characteristics from training data and uses them to create a so-called decision tree. Once the tree is constructed, it can be used to classify data similar to what was used for training. The decision process commences with a single “node” that splits the input data based on one or more characteristics. Subsequently, the data proceeds to one of the nodes in the next layer, depending on the values of the data. At each node, another input feature is chosen, and this recursive process persists until no additional nodes are available. The ultimate node in this decision tree is referred to as a “leaf.” In the case of btagHbb, the leaves contain values within the range of -1 to 1. A value of -1 signifies that the jet likely did not originate from a $b\bar{b}$ pair, whereas a value of 1 suggests a high likelihood that it did. A value of 0 indicates that the algorithm was unable to determine if the jet contained a $b\bar{b}$ pair or not. The btagHbb BDT was trained on simulated $H \rightarrow b\bar{b}$ decays as the signal and QCD multijet events as the background.

Some of the input features which the BDT considered were the impact parameters (or offset between primary and secondary vertices) of tracks in the jet, and the correlation between the directions of the two subjects from the b hadrons. Any dependence on the mass of the particle which decayed to the $b\bar{b}$ pair is avoided to prevent any correlation between the double- b -tag discriminator value and the reconstructed mass of the AK8 jet.

6.2.2 ParticleNet

ParticleNet is an algorithm based on a neural network [69]. Neural networks are algorithms inspired by the human brain and its neurons. A network of artificial neurons with changeable weights, or nodes, is trained on simulated data. The number of layers, as well as the number of nodes per layer, can vary. The first layer can consist of multiple nodes processing different input features simultaneously, and each node in the subsequent layers can be connected to multiple nodes in the previous layer, unlike a BDT. The final layer returns the result. By changing the weight of each node during the training process, it can become better over time. Many jet flavour taggers within CMS have taken advantage of neural networks, with ParticleNet being one of the best-performing ones. ParticleNet is a relatively new tagger and treats each jet as an unordered set of particles, instead of e.g. attempting to identify subjects

and use them as input to the algorithm. It uses various information from jet's Particle Flow (PF) candidates as inputs, such as the differences between the η and ϕ coordinates of the jet axis and each particle, as well as each particle's p_T and energy. Multiple ParticleNet algorithms have been trained to distinguish flavoured jets from the background due to QCD multijets. One of the neural networks is a double- b -tagger. Any dependence on the $b\bar{b}$ pair mass is avoided. The output of the ParticleNet double- b -tagger can be translated to a value between 0 and 1, which is related to the probability that the AK8 jet originated from a $b\bar{b}$ pair.

6.3 Jet Mass Reconstruction

Once a jet has been reconstructed and identified, it is also important to determine the mass of the particle from which the jet originated. This is done using mass reconstruction algorithms which take an AK8 jet as input. Two such techniques are described below.

6.3.1 Softdrop Mass

Softdrop [67] is a jet reconstruction algorithm for estimating the mass of jets with a substructure, such as AK8 jets from $H \rightarrow b\bar{b}$. The Softdrop algorithm aims to remove soft wide-angle radiation, such as pileup, from the jet which could obscure the jet's properties. The Softdrop algorithm declusters the AK8 jet by splitting the jet into two subjets using the so-called Cambridge-Aachen [70] algorithm, and then discards the subjet with the lowest p_T unless the following condition is met:

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^\beta. \quad (6.1)$$

The p_{Ti} is the transverse momentum of subjet i , ΔR_{12} is the angular distance between the two subjets, and R_0 is the radius of the jet (i.e. $R_0 = 0.8$ for an AK8 jet). The two parameters z_{cut} , which removes soft radiation, and β , which removes wide-angle radiation, are the Softdrop parameters. CMS generally uses $z_{\text{cut}} = 0.1$ and $\beta = 0$, thus the Softdrop criteria only depends on the momentum. The Cambridge-Aachen algorithm is originally a jet reconstruction technique, similar to the AK algorithm, except it clusters the spatially closest radiation first. Undoing the last step of this algorithm will result in the two subjets used for the Softdrop criteria. The process of declustering, checking the Softdrop criteria, and removing the softest subjet is repeated until both subjets meet the criteria. The mass is then computed from the remaining jet constituents. If none of the subjets satisfies the Softdrop condition, the mass is set to 0 GeV. This can occur, for example, when the PF objects from one of the b quarks are not contained within the AK8 jet

radius, or if a light jet is incorrectly identified as one of the two selected fat jets. A jet mass close to 0 GeV can also occur if the algorithm removes too many of the PF objects.

6.3.2 ParticleNet

The ParticleNet mass [71] is obtained through the ParticleNet mass regression algorithm that employs a neural network similar to that used in the ParticleNet jet flavour taggers. Again, the input data for the neural network are simple kinematic information of each particle in the AK8 jet, but the output will be the reconstructed mass of the jet. The training for particles decaying to boosted $b\bar{b}$ pairs was performed using simulated events with a flat mass distribution of the decaying particle, hence the algorithm is not dependent on the mass of the original particle. The training was also performed on QCD multijet events, to avoid giving mass to background events. As the neural networks of the double- b -tagger and the mass regression were trained independently, the two outcomes are mass-uncorrelated.

6.4 Signal Model

In the NMSSM, the stable LSP can be singlino-like ($\tilde{\chi}_1^0$) [10]. As the $\tilde{\chi}_1^0$ does not carry any charges, it does not couple directly to SM particles through gauge interaction, and it only has a small coupling to the other SUSY particles because its mixing with the other neutralinos is suppressed [10]. Squarks ($\tilde{q}$) would, therefore, result in decay cascades that end with a $\tilde{\chi}_1^0$ and either a Z, W, or Higgs boson. This study focuses on the case where the final-state particle is a Higgs boson that has the same couplings as the H_{SM} (SM-like) but is lighter than or is the H_{SM} itself, i.e. when $m_{H_1} \leq 125$ GeV [12]. The direct production cross section of the NMSSM light Higgs boson is highly suppressed compared to the H_{SM} , explaining why the particle might have evaded detection previously. The decay cascade of the signal is illustrated in Figure 6.1, where each $\tilde{q}$ decays to a quark and a bino-like Next-to-LSP (NLSP) $\tilde{\chi}_2^0$. The $\tilde{\chi}_2^0$ decays further into a $\tilde{\chi}_1^0$ and a light Higgs boson, where the light Higgs boson decays into a boosted $b\bar{b}$ pair, resulting in an all-hadronic event. It is assumed that other SUSY particles in the NMSSM, not mentioned here, are too heavy to be produced directly at the LHC.

Assuming that the mass of the gluino ($\tilde{g}$) is slightly larger than the squark mass (m_{SUSY}), the gluino would primarily decay to a squark and a quark, $\tilde{g} \rightarrow \tilde{q} + q$, before following the same squark decay as described earlier. A small mass difference between the gluino and the squark results in similar decay and final-state particles for both the squark and gluino production modes, as only a small amount of the total momentum is carried away by the additional quark produced in the gluino production mode. The cross section of producing squark or gluino pairs ($\tilde{q}\tilde{q}$, $\tilde{q}\tilde{g}$, $\tilde{g}\tilde{g}$) with mass m_{SUSY} in proton-proton collisions

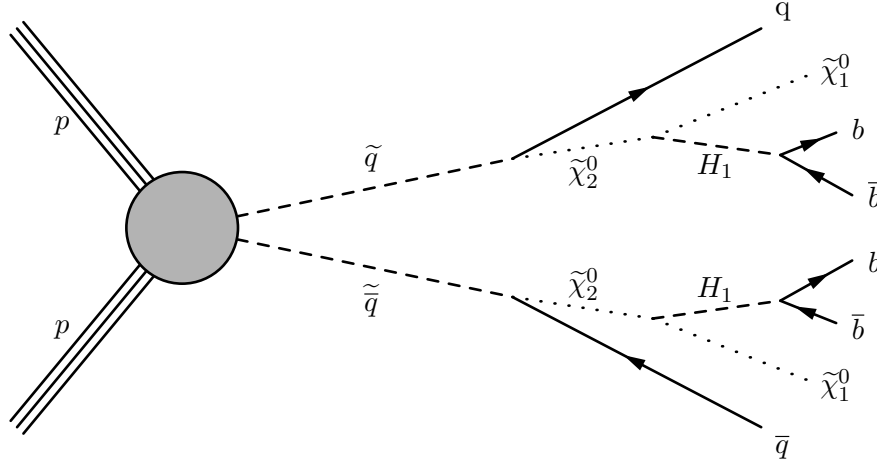


Figure 6.1: A Feynman diagram of the $\tilde{q}\tilde{q}$ signal model [12]. The $\tilde{q}\tilde{q}$ and the $\tilde{g}\tilde{g}$ signal models follow the same characteristics.

at 13 TeV can be seen in Table 6.1. For smaller values of m_{SUSY} , $m_{\text{SUSY}} < 1200$ GeV, the resulting H_1 are less boosted and require different techniques for reconstruction compared to boosted jets, hence are not considered. In the published analysis, values of m_{SUSY} up to about 2500 GeV were excluded. In anticipation of the larger event sample in Run 3 and improved analysis techniques, values of m_{SUSY} up to 3200 GeV are considered, corresponding to a signal production cross section about a factor of 10 smaller than previously excluded.

Table 6.1: Squark/gluon pair-production cross sections calculated using NNLL-FAST 1.1 [72] for squark masses set to m_{SUSY} and gluino masses set to 1% larger than m_{SUSY} .

$m_{\text{SUSY}}[\text{GeV}]$	$\sigma(pp \rightarrow \tilde{q}\tilde{q}, \tilde{q}\tilde{g}, \tilde{g}\tilde{g}) [\text{fb}]$	Uncertainty
1200	582.3	8%
1400	194.1	9%
1600	69.71	9%
1800	26.29	10%
2000	10.29	11%
2200	4.114	13%
2400	1.668	14%
2600	0.6792	16%
2800	0.2765	18%
3000	0.1117	21%
3200	0.04512	24%

The light squark flavours, i.e. excluding sbottoms and stops, are assumed to have the same mass m_{SUSY} . The sbottom and stop masses are assumed heavier than accessible via direct production. Apart from m_{SUSY} and m_{H_1} ,

the signal model is also dependent on the unknown masses of $m_{\chi_2^0}$, and $m_{\chi_1^0}$. The two additional degrees of freedom can be rewritten as

$$R_M \equiv \frac{m_{H_1}}{m_{\chi_2^0}}, \quad \Delta_M \equiv m_{\chi_2^0} - m_{H_1} - m_{\chi_1^0}. \quad (6.2)$$

To get the p_T^{miss} -suppressed signal, choosing $R_M = 0.99$ and $\Delta_M = 0.1$ GeV will yield a light $\tilde{\chi}_1^0$ which carries very little momentum compared to the H_1 . Values of $R_M > 0.9$ were shown to yield a similar signal, however, for lower R_M values the p_T^{miss} is no longer negligible and standard SUSY searches that include p_T^{miss} would be more appropriate [12]. As a consequence of the chosen parameters, the squark and gluino decay cascade would end in a soft $\tilde{\chi}_1^0$ and a boosted H_1 . The majority of the time, the H_1 decays into a boosted $b\bar{b}$ pair, as seen for the branching fractions in Table 6.2, because the $H_1 \rightarrow b\bar{b}$ branching fractions are expected to be SM-like [10]. The branching fractions were calculated using HDECAY [73, 74], assuming a H_{SM} but changing the mass. Both H_1 bosons need to decay to $b\bar{b}$ pairs for it to be considered a signal event. With these chosen values of R_M and Δ_M , the branching fractions of the decays $\tilde{g} \rightarrow \tilde{q} + q$ and $\tilde{\chi}_2^0 \rightarrow H_1 + \tilde{\chi}_1^0$ are assumed to be unity.

Table 6.2: Branching fractions for $H_1 \rightarrow b\bar{b}$ for various Higgs masses. The branching fractions were calculated using HDECAY [73, 74].

m_{H_1} [GeV]	Branching Fraction ($H_1 \rightarrow b\bar{b}$)
30	0.868
35	0.867
40	0.865
50	0.858
60	0.850
70	0.840
80	0.829
90	0.816
100	0.795
110	0.749
125	0.581

6.5 Event simulation

To simulate and predict the outcomes of the data analysis, MC simulation samples are used to develop and test the signal event selection, as well as for making background estimations. To generate the NMSSM signal MC samples, multiple samples were produced by varying the values of m_{H_1} and m_{SUSY} within the mass ranges of 30-125 GeV and 800-3200 GeV, respectively. The masses of the other particles predicted by the NMSSM, such as the charginos,

were set to infinity as they were assumed to be out of reach at the LHC. The dominant background for the signals under consideration arises from QCD multijet events. However, the $t\bar{t}$ background must also be taken into account since the top quark decays nearly 100% of the time into a bottom quark and a W boson [75]: the presence of a b quark among other merged decay products of a top quark² means such massive fat jets are more likely to mimic a signal $H \rightarrow b\bar{b}$ jet. The Z +jets and W +jets processes can also generate background events via decays into quark anti-quark pairs, particularly $Z \rightarrow b\bar{b}$ where at least one jet is likely to get double- b -tagged.

The official CMS MC samples were used for this study, which were generated using MADGRAPH5_aMC@NLO [76, 77], POWHEG [78–80], and PYTHIA [81], while GEANT4 [82] was used for simulating the particle interactions with the CMS detector. These generators and their corresponding settings have been validated to give the best description of CMS data so far. Both the QCD and $t\bar{t}$ samples were produced in bins of H_T . The simulations were then stored in the NanoAOD [83] file format, which is a compact data event file format used in CMS.

6.6 NanoAODTools

The analysis framework used in this study is based on NANOAODTOOLS [84, 85], which is a common Python-based CMS analysis tool. The framework incorporates tools to evaluate commonly used systematic uncertainties and provides corrections based on scale factors with calibration values recommended by CMS, which are applicable for Run 2 (2016-2018) data. These include corrections and systematic uncertainties for AK4 and AK8 jet energies and mass reconstructions, as well as pileup and scale factors for b -taggers to account for any discrepancies in b -tagging performance between data and simulated events. However, since ParticleNet is a relatively new double- b -tag discriminator, manual implementation of scale factors and uncertainties specific to ParticleNet is required (Section 6.9).

6.7 Event Selection

To separate the signal events from background, various quantities of the events are investigated. As most of the energy from the squarks and gluons is transformed into jets, the signal events should have a high H_T . Therefore, high H_T events are selected at the trigger level. In this chapter, only the data from the year 2018 of Run 2 were used for the studies, which corresponds to an integrated luminosity of $59.7 \pm 1.5 \text{ fb}^{-1}$ before any selections were made. Only

²Unlike the lighter quarks, the top quark’s lifetime is too short for hadronisation to occur.

the events that passed a specific HLT H_T trigger were considered for selecting signal events. The published analysis was used as a baseline, where a minimum H_T trigger threshold of 1050 GeV was used due to being the lowest available H_T trigger for Run 2. This threshold removed not far from 100% of the QCD events and 92% of the $t\bar{t}$ events. Although the main backgrounds for the NMSSM signal model come from QCD and $t\bar{t}$, there are also minor contributions from W +jets and Z +jets backgrounds. The trigger selection rejected the majority of the W +jets and Z +jets events. The percentage of signal events rejected depended slightly on m_{SUSY} and m_{H_1} , however, $< 1\%$ of the events were removed in general. Note that after this trigger requirement, the contributions from the QCD events remain much larger than for any other process, thus it is still the dominant background.

The purpose of this study was to study the performance of ParticleNet and determine if it would be beneficial for the analysis to change from btagHbb and Softdrop mass to the ParticleNet double- b -tagger and ParticleNet mass regression. Therefore, all analysis cuts from the published analysis were retained when using btagHbb and Softdrop mass as the cuts had already been optimised for signal sensitivity [12]. This also provided the opportunity to cross-check any results with what had been studied previously. When using ParticleNet, the cuts related to the double- b -tag discriminator were chosen such that the event count for MC signal events was roughly the same as for btagHbb and Softdrop after all cuts had been performed. This was done in order to easily compare the amount of background events remaining after the event selections. The set of algorithms with the fewest background events remaining has the better performance. It is worth noting that further studies on which cuts to use for ParticleNet can be conducted to achieve optimal performance, particularly if the ParticleNet performance is showing good potential at the end of this investigation. However, such optimisations are out of scope of this thesis.

Examples using signal events presented here will employ a signal sample with $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV unless otherwise noted.

6.7.1 Kinematic Event Selection

The signal event selections were, in part, determined by the kinematics of the event, including the H_T and momenta of the jets within the event. For the kinematic event selection in the published analysis, a minimum H_T of 1500 GeV was chosen, which is sufficiently above the HLT 1050 GeV H_T trigger such that almost no relevant events are lost due to any inefficiencies in the trigger. The H_T distribution of various signal and background samples is shown in Figure 6.2. The H_T was calculated as the scalar sum of all AK4 jets with a $p_T > 40$ GeV and $|\eta| < 2.4$. The $|\eta| < 2.4$ cut was used to avoid efficiency losses in the ECAL endcaps due to radiation damage for the 2018 Run 2 dataset [86]. The selected events can also be split into H_T categories

to improve sensitivity for the larger values of m_{SUSY} as the signal event H_T increases with m_{SUSY} , while the H_T for backgrounds remains unchanged. A high H_T bin would contain much of the signal event, while most of the background events end up in the lower H_T bins. As the H_T bins should have minimal impact on the performance comparisons between btagHbb/Softdrop and ParticleNet, only one H_T bin was used in this study.

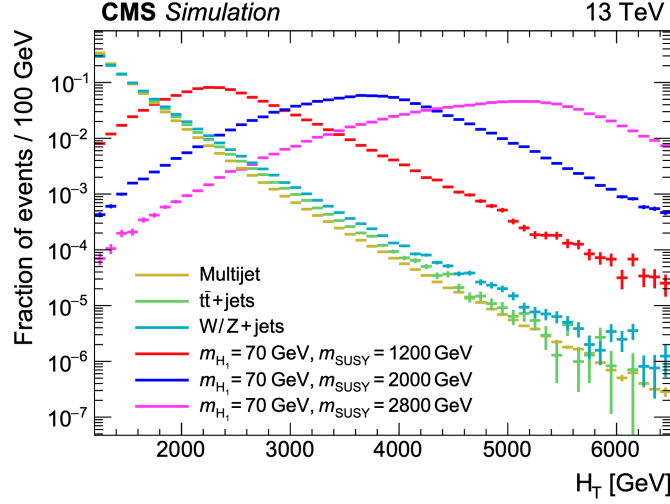


Figure 6.2: The H_T distribution for various background and signal events [12]. All events satisfy the H_T trigger selection.

The $b\bar{b}$ pairs resulting from the decay of the boosted light Higgs bosons were reconstructed as AK8 jets, also known as fat jets. The two AK8 jets with the highest double- b -tag discriminator value were selected to represent the $b\bar{b}$ pairs. For the $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV signal sample, both jets were correctly selected 63% of the time for btagHbb and 88% of the time for ParticleNet. Similar values were found for the other signal samples, hence the ParticleNet double- b -tagger is better at selecting the correct signal jets. The two selected fat jets were randomly assigned the names *FatJetA* and *FatJetB* to ensure symmetrical distributions of the jets. The quarks from the quark decay were represented by smaller AK4 jets. The following requirements from the published analysis were used for selecting the signal events after choosing the fat jets:

- $H_T > 1500$ GeV.
- The two selected AK8 jets must have $p_T > 300$ GeV and $|\eta| < 2.4$, corresponding to the two boosted $b\bar{b}$ pairs.
- There must be at least one AK4 jet with $p_T > 300$ GeV and $|\eta| < 2.4$, corresponding to a quark from one of the squark decays. The AK4 jet

must also be separated from both AK8 jets by a distance $\Delta R > 1.4$ (see Equation 2.5) so it does not contain any of the same PF candidates as the AK8 jets.

The choice of a jet $p_T > 300$ GeV threshold was determined from the p_T distribution of the signal AK8 and AK4 jets [87]. For the background MC samples after the trigger selection, the kinematic event selection removed 95% and 90% of the remaining QCD and $t\bar{t}$ events, respectively. About 93% of the remaining W +jets and Z +jets events were also rejected. For the $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV signal sample, 17% of the events were rejected.

6.7.2 Double- b -tag Selection

The double- b -tag discriminators of each of the two AK8 jets together form a 2D space. Since jets from $b\bar{b}$ decays typically have a high discriminator value per definition, they are expected to cluster in the upper right corner of the parameter space. Hence, a Tag Region (TR) for selecting signal events is defined as the area where the sum of the double- b -tag discriminator values exceeds a certain threshold. In the previous Run 2 analysis, a cut of 1.3 was applied for the sum of the btagHbb discriminators, illustrated in Figure 6.3. Two additional regions were defined in this double- b -tag distribution to estimate the background from data (Section 6.8): the Control Region (CR) and the Validation Region (VR). The CR has a low double- b -tag discriminator value for both selected AK8 jets and should predominantly contain QCD background events, and is used to perform background estimations. The VR is used to validate the background estimations.

When using the btagHbb discriminator after the kinematic event selection, 64% of the events fall within the btagHbb TR for the $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV signal sample. For very low or high H_1 masses, the fraction of events goes down to around 50%. To make performance comparisons between the taggers, a cut on the double- b -tag discriminator sum of 1.95 was selected as a starting point for ParticleNet, corresponding to roughly a similar event count for signal MC within the TR as for btagHbb³. It is worth noting that the discriminator values range from -1 to 1 for btagHbb and from 0 to 1 for ParticleNet. The CR and VR for ParticleNet were chosen to cover the same fraction of area as for btagHbb with further optimisation left for future work. The definition of the CR and VR do not impact the comparison between btagHbb and ParticleNet, as they are only used for the data-driven

³For a fairer comparison of background yields, it would have been better to apply a tighter cut that yields exactly the same number of events after the mass event selection in Section 6.7.3. This would correspond to a double- b -tag sum cut of 1.99. However, choosing a tighter cut could risk too few QCD background events for the background studies for ParticleNet. Hence, the number of signal events after all selections is currently a factor ~ 2 more for ParticleNet compared to btagHbb. The cut will be optimised for discovery significance in the future.

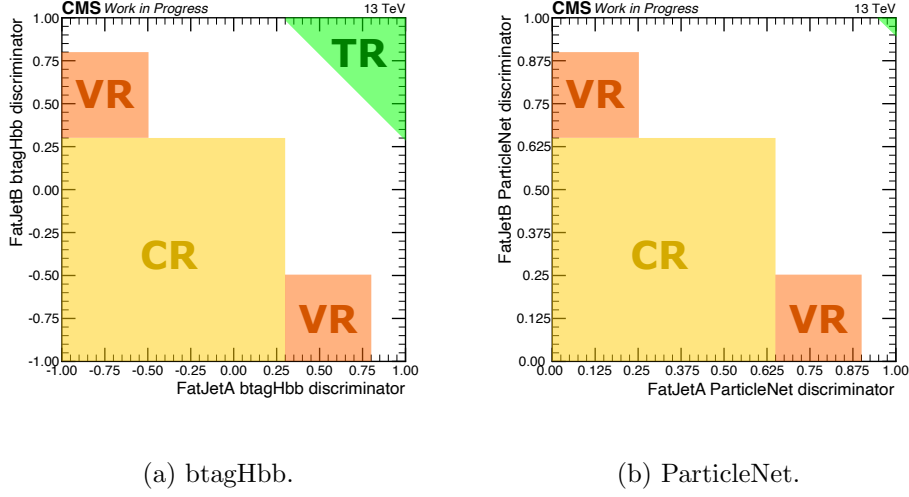


Figure 6.3: The double- b -tag region definitions for the btagHbb (left) and ParticleNet (right) taggers. The TR (green) contains all events that passed the signal selection. The CR (red) and the VR (yellow) are used for background estimations.

background estimation (Section 6.8), where the requirements of CR is to include a larger fraction of background events. The requirement for VR is to have a comparable, or larger, event count than the TR. Both these requirements are satisfied for ParticleNet as well when using the region definitions optimised for btagHbb.

Figure 6.4 illustrates the 2D distribution of the double- b -tag discriminator values of a signal sample for both btagHbb and ParticleNet. As expected, most signal events end up in the upper right corner. The differences in the distribution shape are due to the distinct definitions of the double- b -taggers. ParticleNet is related to the probability that the fat jet contained two b quarks, whereas btagHbb returns a value close to 1 if the jet is believed to contain two b quarks and -1 if it is not. If the decision is inconclusive, a value near 0 is returned.

Figures 6.5-6.8 show the 2D double- b -tag distributions for the QCD, $t\bar{t}$, W +jets, and Z +jets background samples. The background events are spread out over the whole area, with ParticleNet having a stronger preference for small discriminator values than btagHbb. The spread of the signal events being smaller for ParticleNet, and the QCD events being concentrated to the bottom left, suggests that ParticleNet is better at separating the signal from the background. The double- b -tagging performance is investigated further in Section 6.10. The $t\bar{t}$ background has a larger fraction of events with higher double- b -tag values due to b quarks arising from top quark decays: for btagHbb (ParticleNet), the TR contains 4% ($< 0.01\%$) and 15% (0.4%) of the remaining

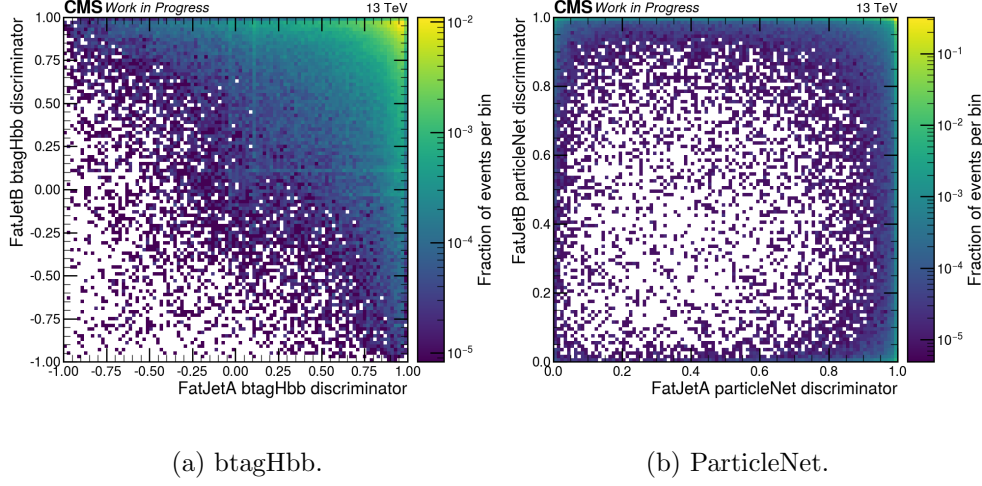


Figure 6.4: Distribution of the double- b -tag discriminator values for the two selected AK8 jets for a signal sample with $m_{H_1} = 80 \text{ GeV}$ and $m_{\text{SUSY}} = 2800 \text{ GeV}$, for the btagHbb algorithm (left) and the ParticleNet algorithm (right).

events after the kinematic event selection for QCD and $t\bar{t}$, respectively. The double- b -tag selection for the btagHbb algorithm rejected 96% and 93% of the events remaining after the kinematic event selection for W +jets and Z +jets, respectively. The selection using ParticleNet removed 99.9% of the events for both W +jets and Z +jets. While comparing the distribution of Z +jets events obtained using ParticleNet with its counterpart from btagHbb (Figure 6.8), it's evident that the ParticleNet distribution exhibits a more signal-like profile. This improvement underscores ParticleNet's ability to enhance the identification of $b\bar{b}$ pairs from the Z boson decays.

The horizontal and vertical lines near zero observed for btagHbb in Figures 6.5-6.8, represent instances where the double- b -tagger cannot confidently determine the presence or absence of a $b\bar{b}$ pair within the fat jet. This phenomenon is more pronounced when the selected AK8 exhibits low p_T and is observable in both actual data and MC simulated events [87]. Importantly, since these lines are situated significantly away from the TR, they have no impact on the analysis.

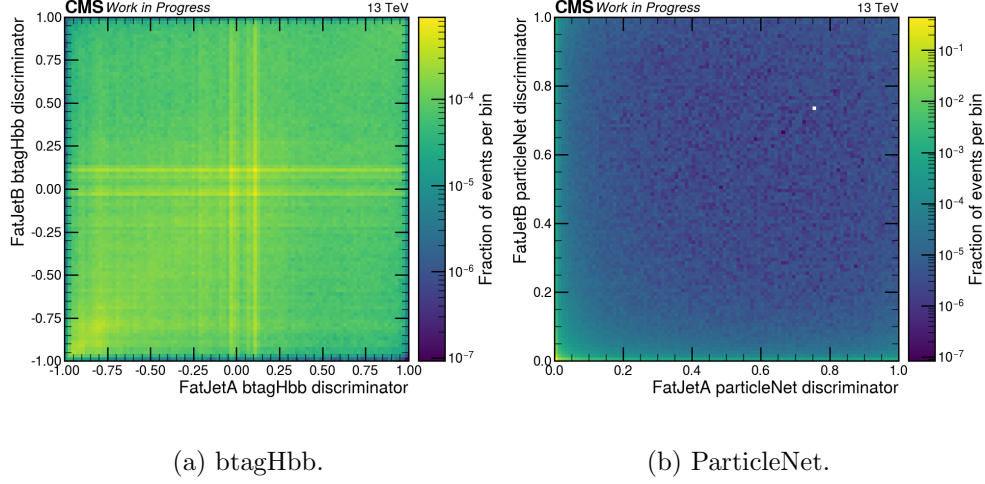


Figure 6.5: Distribution of the double- b -tag discriminator values for the two selected AK8 jets for QCD background events, for the btagHbb algorithm (left) and the ParticleNet algorithm (right). The white bin in the ParticleNet distribution is due to a limited number of events in the MC sample.

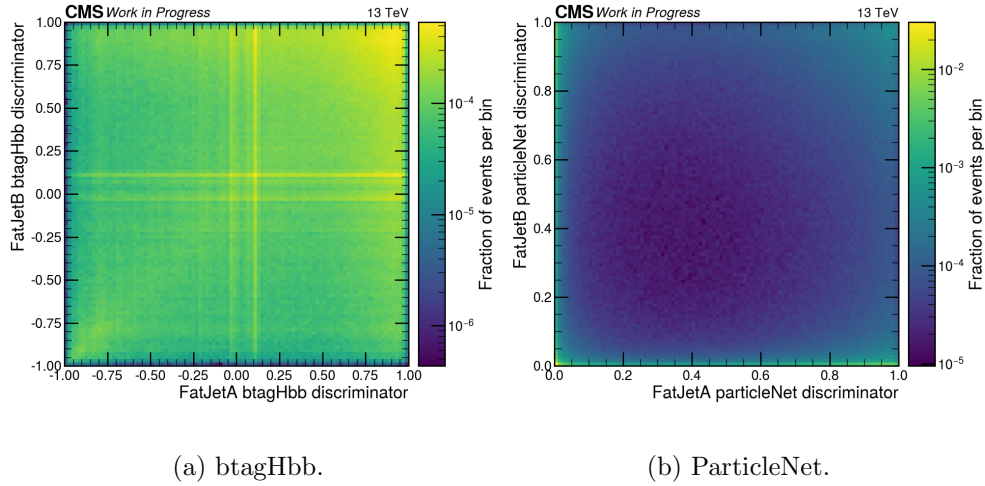


Figure 6.6: Distribution of the double- b -tag discriminator values for the two selected AK8 jets for $t\bar{t}$ background events, for the btagHbb algorithm (left) and the ParticleNet algorithm (right).

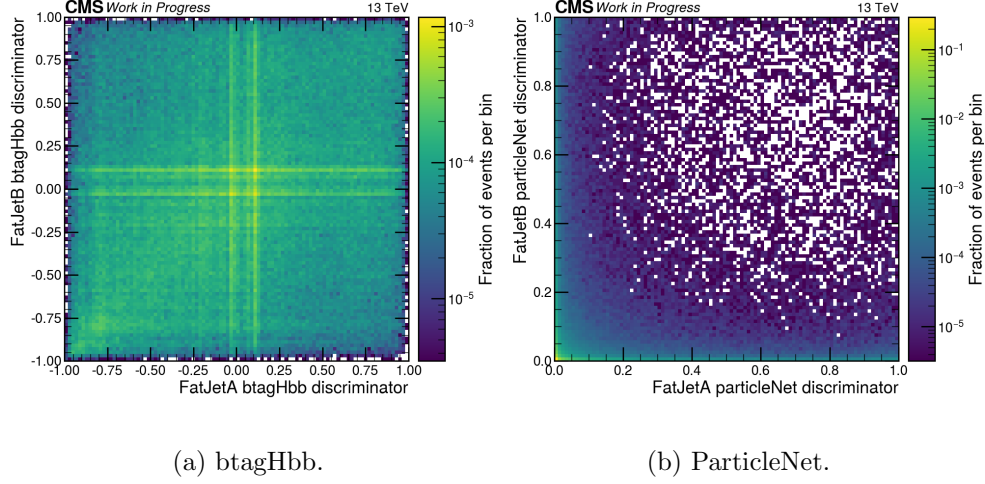


Figure 6.7: Distribution of the b -tag discriminator values for the two selected AK8 jets for W +jets background events, for the btagHbb algorithm (left) and the ParticleNet algorithm (right).

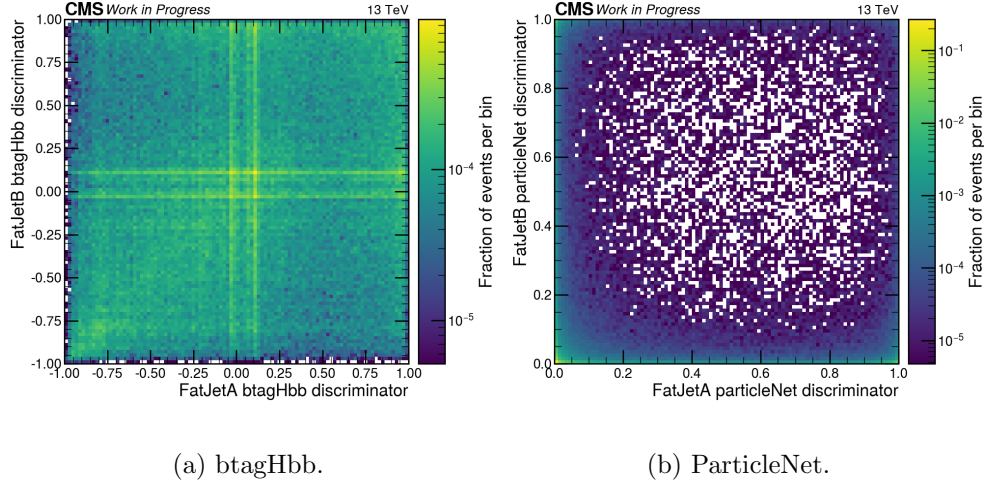


Figure 6.8: Distribution of the b -tag discriminator values for the two selected AK8 jets for Z +jets background events, for the btagHbb algorithm (left) and the ParticleNet algorithm (right).

6.7.3 Mass Event Selection

Event selection based on the masses of the AK8 jets is also performed, as the masses of the two selected jets should match if they both originated from the same type of particle. This requires accurate reconstructions of the jet mass. Two different mass evaluations are used for the two double- b -tag discriminators: Softdrop mass [67] and ParticleNet mass [71] for btagHbb and ParticleNet, respectively.

A comparison of the mass distribution of signal events is shown in Figure 6.9. The performance of the ParticleNet mass is superior to the Softdrop mass as more events end up close to the true value of m_{H_1} , i.e. it does not suffer as much from the issue of massless jets. The spread of the main peak relative to its height is also less for ParticleNet, showing that the ParticleNet mass resolution is better even for jets with two subjets that can be successfully identified by the Softdrop algorithm. The fat jets used in Figure 6.9 are the two AK8 jets with the highest double- b -tag values, thus the fact that ParticleNet chooses the correct fat jets more frequently than btagHbb might also affect the spread and the peak around zero.

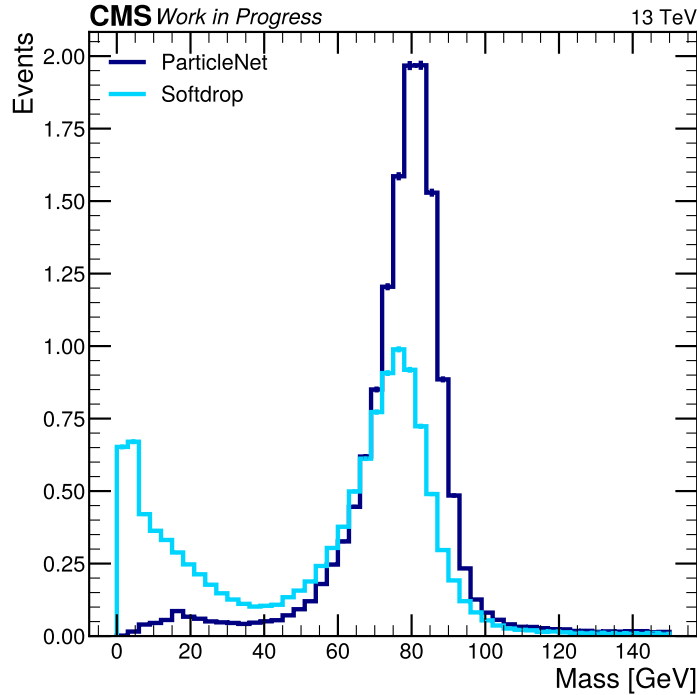


Figure 6.9: A comparison of the reconstructed AK8 jet masses using the Softdrop algorithm and the ParticleNet mass regression for a signal sample ($m_{H_1} = 80$ GeV, $m_{\text{SUSY}} = 2800$ GeV).

As both selected AK8 jets are expected to originate from a light Higgs boson, their mass should be close to that of m_{H_1} . In the 2D space defined by the mass of the two AK8 jets, the signal should therefore appear along a diagonal, determined by m_{H_1} . In the published analysis, multiple signal regions are defined along this diagonal to accommodate a range of masses. As is typical for physics analysis, these regions were optimised to maximise the signal significance when using btagHbb. Additionally, sideband regions are defined next to the signal mass regions for use in data-driven background estimations, as discussed in Section 6.8. An example of the signal and sideband regions which were used is shown in Figure 6.10. The same signal and sideband mass regions were applied to ParticleNet as a baseline for this study. A discussion on optimising the mass region grid for ParticleNet is found in Section 6.11.2.

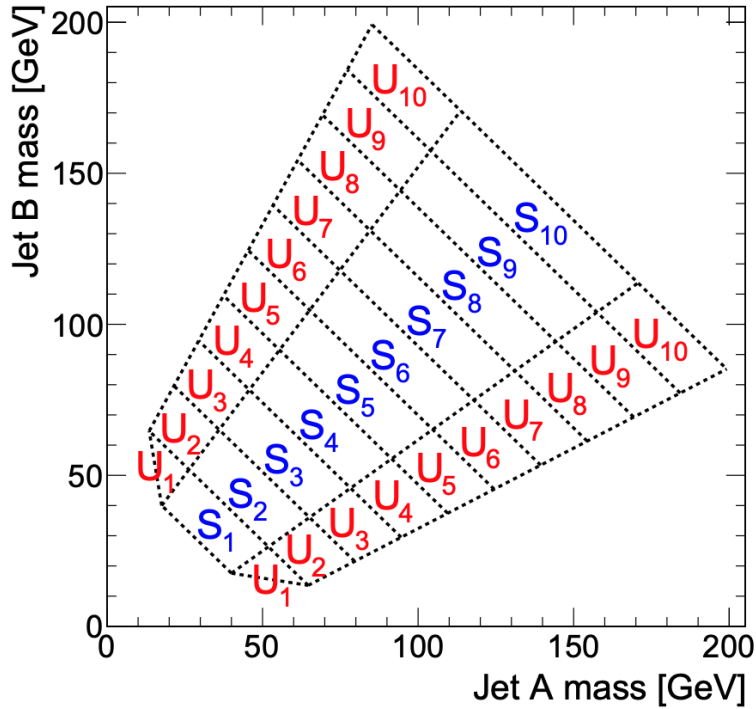


Figure 6.10: Example of mass regions in the 2D jet mass plane developed for the btagHbb double- b -tagger and the Softdrop mass [12]. The signal mass regions are denoted S_i , and the sideband mass regions are denoted U_i . This mass grid was used for both the Softdrop and ParticleNet mass.

The 2D mass distributions for both signal and background (QCD, $t\bar{t}$, and W +jets and Z +jets) are presented in the following sections, along with a comparison of the mass distributions before and after the double- b -tag selection. This allows for a better visual understanding of the performance of the two double- b -tag discriminators.

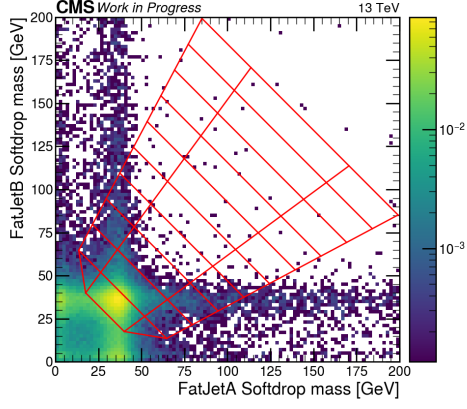
Signal Sample

The mass distributions of the selected AK8 jets after the double- b -tag selection are shown in Figures 6.11-6.12. Applying the double- b -tag selection removes a lot of the events where only one of the two jets really came from $H \rightarrow b\bar{b}$, especially for ParticleNet. This can be seen from the linear features in the plot where at a fixed value of mass for one of the AK8 jet, there is a spread of the mass values for the other jet because it was either not the correct jet or the mass was incorrectly reconstructed. The mass distribution prior to the double- b -tag selection can be seen in Appendix C.1. From the figures, it can be seen that the AK8 mass does indeed increase with m_{H_1} , and that m_{SUSY} has less of an impact on the mass distributions. The distributions using ParticleNet contain a larger fraction of events along the diagonal where the signal mass regions are located, as the Softdrop algorithm condition is quite often too harsh, resulting in 0 GeV masses, thus deemed useless for our analysis.

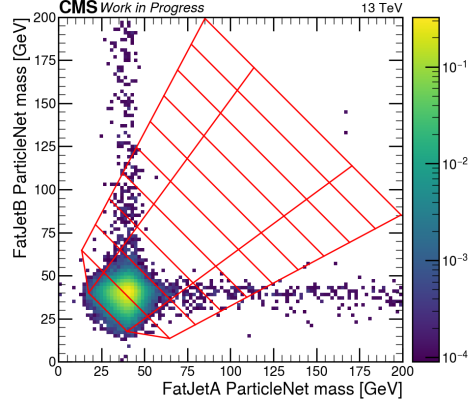
After applying all the selection criteria, 49% of the remaining events were observed in any of the signal mass regions for the Softdrop mass combined with btagHbb, and 98% of the events were observed for the ParticleNet mass and its corresponding double- b -tagger, for the signal with $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV. This corresponds to 24% and 42% of the total number of signal events before any selections were performed for btagHbb and ParticleNet, respectively.

QCD Sample

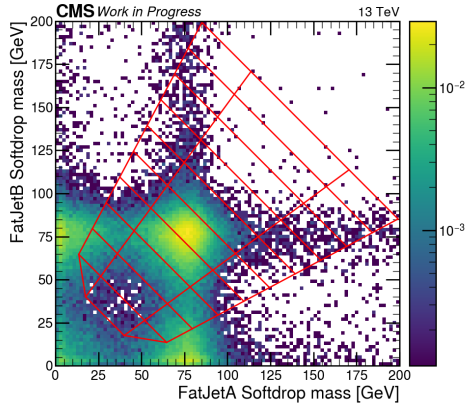
Figure 6.13 shows the 2D distribution of the two AK8 jets prior to the double- b -tag selection for the QCD background. Since the QCD background does not originate from a decaying massive particle, the jet mass from this background is typically small, and the number of background events decreases as the jet mass increases. This is true for both Softdrop and ParticleNet mass. The QCD data-driven background estimation assumes a uniform mass distribution along a signal region and its corresponding sideband region for the QCD background, and this is still valid after switching to ParticleNet mass, apart from slight peaks which can be observed at approximately 15 GeV. These lines are a result of the ParticleNet mass regression algorithm not being trained for signal mass values below this threshold. The QCD samples which were used for training did not have this threshold, which causes the lines to be smooth and not hard thresholds. A small adjustment of the mass grid could avoid any such contribution in the sideband regions, though future retraining of ParticleNet for Run 3 including lower-mass signal events is expected to remove the 15 GeV feature.



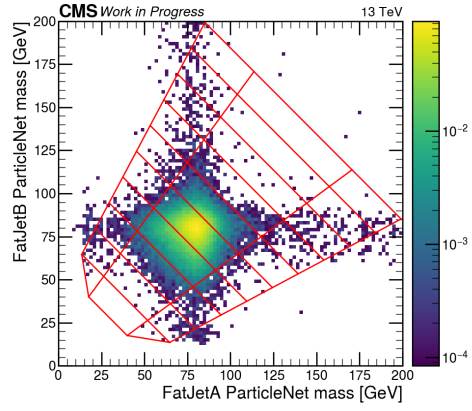
(a) Softdrop, signal sample with $m_{H_1} = 40$ GeV, $m_{\text{SUSY}} = 2800$ GeV.



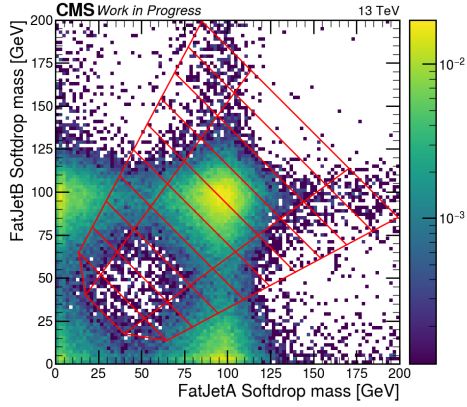
(b) ParticleNet, signal sample with $m_{H_1} = 40$ GeV, $m_{\text{SUSY}} = 2800$ GeV.



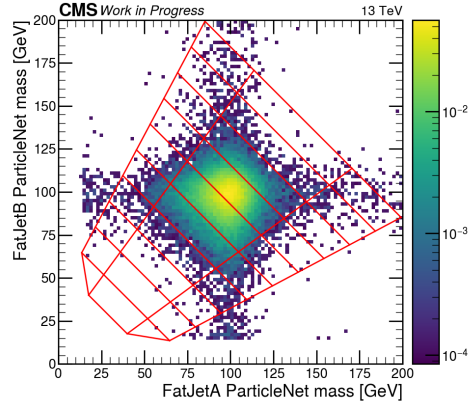
(c) Softdrop, signal sample with $m_{H_1} = 80$ GeV, $m_{\text{SUSY}} = 2800$ GeV.



(d) ParticleNet, signal sample with $m_{H_1} = 80$ GeV, $m_{\text{SUSY}} = 2800$ GeV.

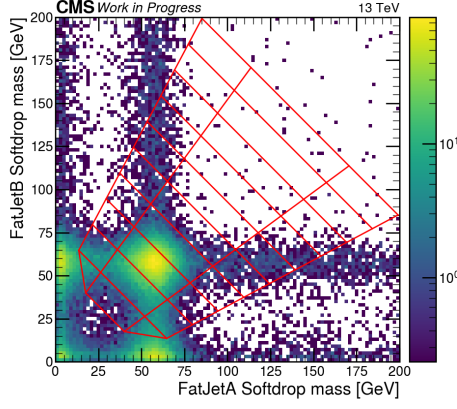


(e) Softdrop, signal sample with $m_{H_1} = 100$ GeV, $m_{\text{SUSY}} = 2800$ GeV.

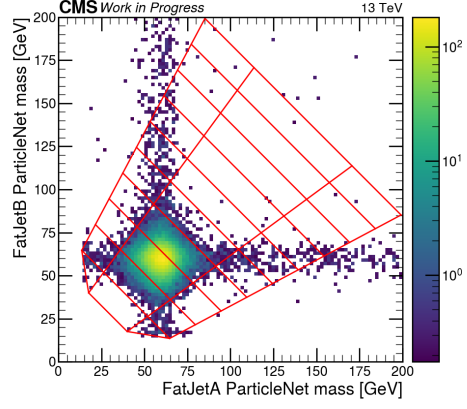


(f) ParticleNet, signal sample with $m_{H_1} = 100$ GeV, $m_{\text{SUSY}} = 2800$ GeV.

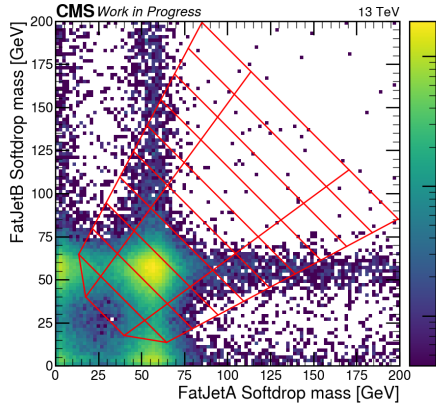
Figure 6.11: Distribution of the reconstructed jet mass for the two selected AK8 jets for various m_{H_1} masses for the Softdrop algorithm with btagHbb (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection and the double- b -tag selection were included.



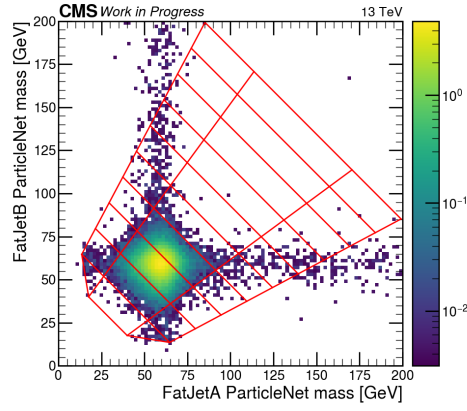
(a) Softdrop, signal sample with $m_{H_1} = 60 \text{ GeV}$, $m_{\text{SUSY}} = 1200 \text{ GeV}$.



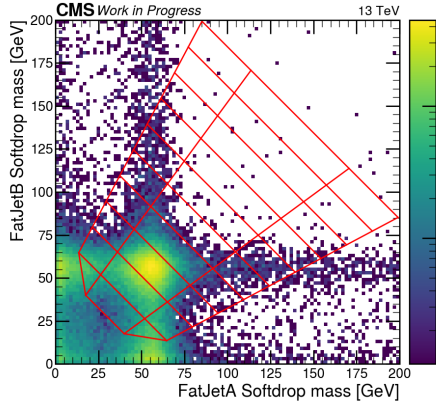
(b) ParticleNet, signal sample with $m_{H_1} = 60 \text{ GeV}$, $m_{\text{SUSY}} = 1200 \text{ GeV}$.



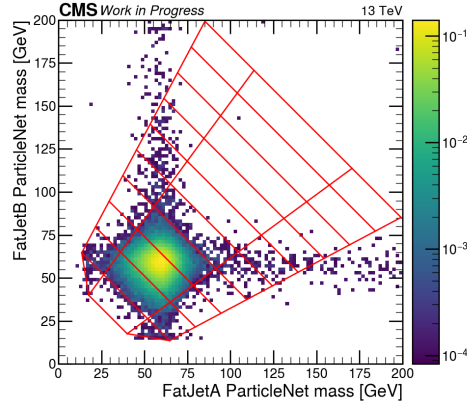
(c) Softdrop, signal sample with $m_{H_1} = 60 \text{ GeV}$, $m_{\text{SUSY}} = 2000 \text{ GeV}$.



(d) ParticleNet, signal sample with $m_{H_1} = 60 \text{ GeV}$, $m_{\text{SUSY}} = 2000 \text{ GeV}$.



(e) Softdrop, signal sample with $m_{H_1} = 60 \text{ GeV}$, $m_{\text{SUSY}} = 2800 \text{ GeV}$.



(f) ParticleNet, signal sample with $m_{H_1} = 60 \text{ GeV}$, $m_{\text{SUSY}} = 2800 \text{ GeV}$.

Figure 6.12: Distribution of the reconstructed jet mass for the two selected AK8 jets for various m_{SUSY} masses for the Softdrop algorithm with btagHbb (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection and the double- b -tag selection were included.

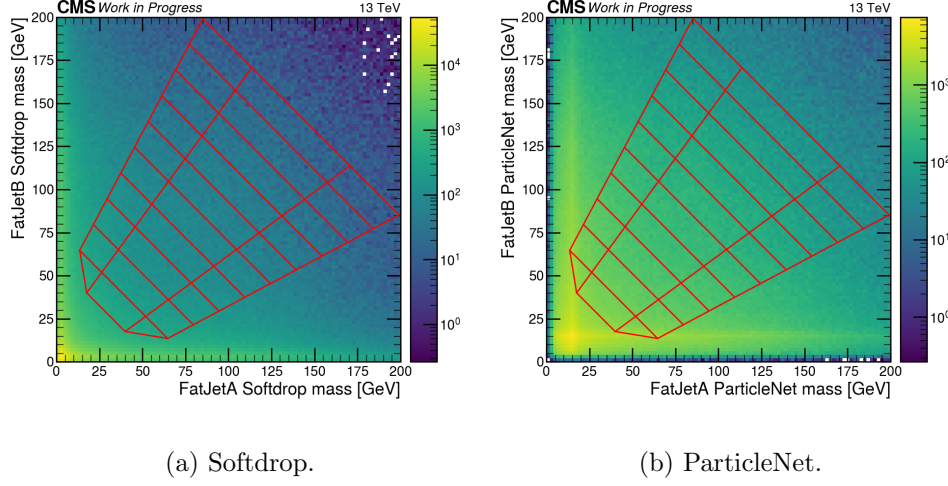


Figure 6.13: Distribution of the reconstructed jet mass for the two selected AK8 jets for the QCD background for the Softdrop algorithm (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection were included.

Despite the kinematic and double- b -tag event selections, some background events still remain. However, as shown in Figure 6.14, the ParticleNet tagger was able to remove the majority of the QCD background. For the $b\text{tagHbb}$ discriminator and the Softdrop mass, 5% of the remaining events were located in any of the signal mass regions, while for the ParticleNet method, it was 27% of the remaining events. However, the final number of QCD events that passed all selections was a factor of 100 smaller for ParticleNet.

$t\bar{t}$ Sample

Figure 6.15 displays the 2D mass distribution of the $t\bar{t}$ background prior to the double- b -tag selection. The selected AK8 jets from $t\bar{t}$ may contain all or some of the top quark decay products such as b -quarks, thus their mass distribution does not peak near zero. Instead, multiple peaks are visible because of the decays of W bosons (~ 80 GeV) and top quarks (~ 172 GeV), along with the 15 GeV ParticleNet mass artefact mentioned earlier.

The mass distribution after the double- b -tag selection, shown in Figure 6.16, reveals that the ParticleNet tagger outperforms $b\text{tagHbb}$ in rejecting the $t\bar{t}$ background: 20% of the remaining events were detected in the signal region for $b\text{tagHbb}$ and Softdrop, while 25% of the remaining events were detected for ParticleNet. A factor of 10 fewer $t\bar{t}$ events passed all selections for ParticleNet compared to $b\text{tagHbb}$, although the enhancement is not as significant as it was for the QCD background

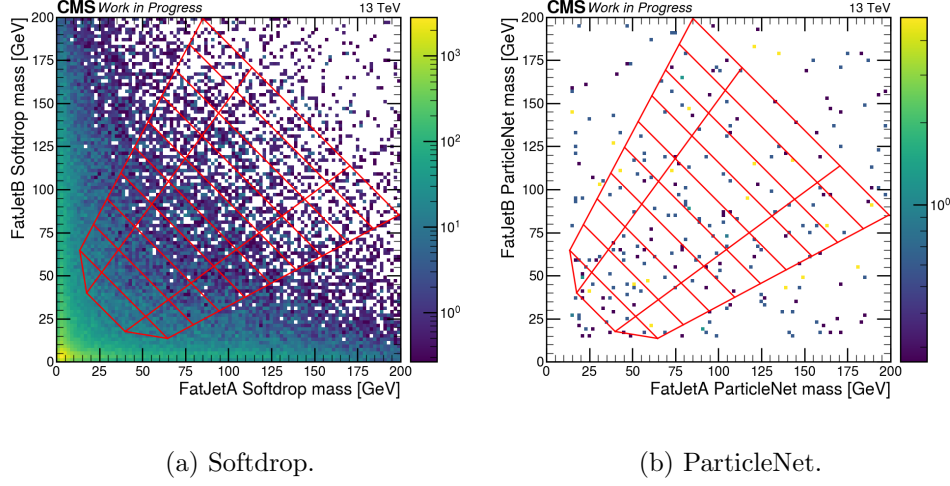


Figure 6.14: Distribution of the reconstructed jet mass for the two selected AK8 jets for the QCD background for the Softdrop algorithm with btagHbb (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection and the double- b -tag selection were included.

W +Jets and Z +Jets Samples

The mass distributions of the selected fat jets prior to the double- b -tag selection can be seen in Figure 6.17 (Figure 6.18) for the W +jets (Z +jets) background. Excesses occur at the W and Z boson masses (~ 80 GeV and ~ 90 GeV, respectively) due to their similarities with the signal as they can decay into b quarks. Again, the 15 GeV ParticleNet mass artefact is observed in the mass distributions.

The mass distribution after the double- b -tag selection is shown in Figures 6.19-6.20. The signal mass regions removed around 94% and 92% of the remaining events for W +jets and Z +jets, respectively, when using the Softdrop mass combined with the btagHbb double- b -tagger. The fraction of events after the double- b -tag selection removed for ParticleNet were 71% and 59% for W +jets and Z +jets, respectively.

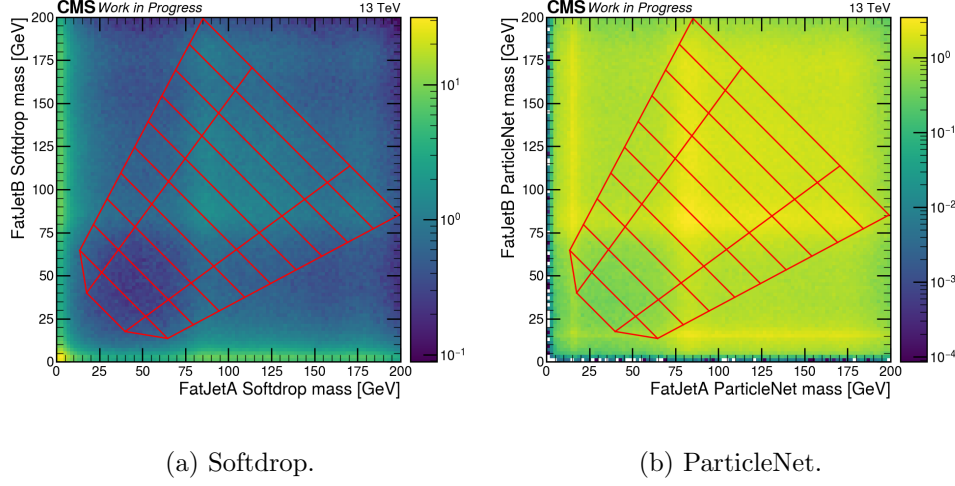


Figure 6.15: Distribution of the reconstructed jet mass for the two selected AK8 jets for the $t\bar{t}$ background for the Softdrop algorithm (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection were included.

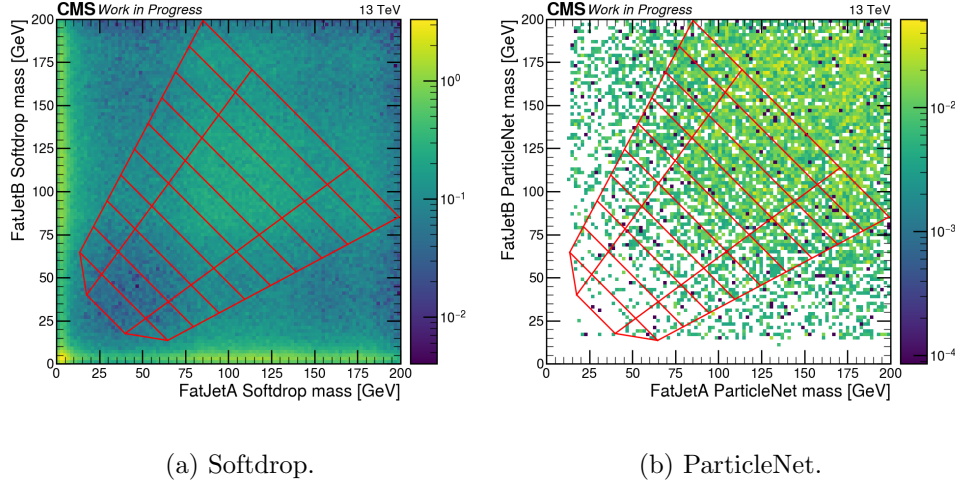


Figure 6.16: Distribution of the reconstructed jet mass for the two selected AK8 jets for the $t\bar{t}$ background for the Softdrop algorithm with btagHbb (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection and the double- b -tag selection were included.

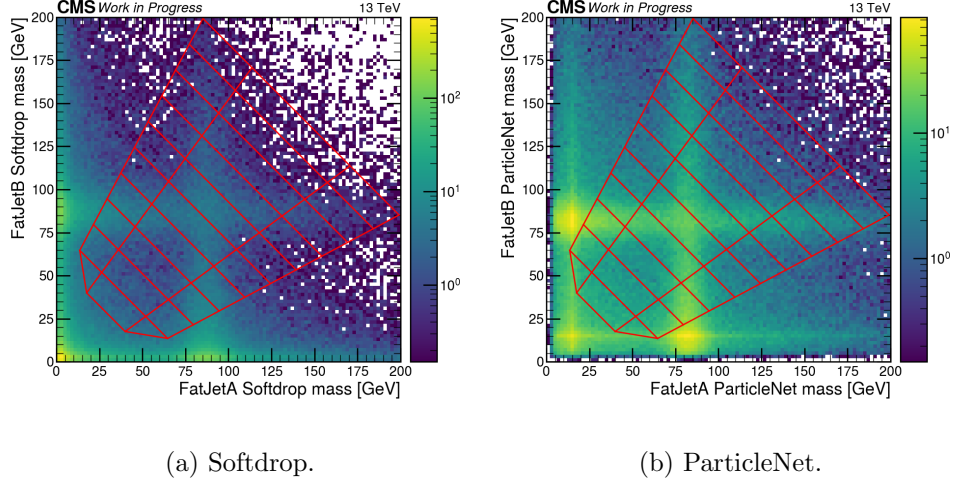


Figure 6.17: Distribution of the reconstructed jet mass for the two selected AK8 jets for the W +jets background, for the Softdrop algorithm (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection were included.

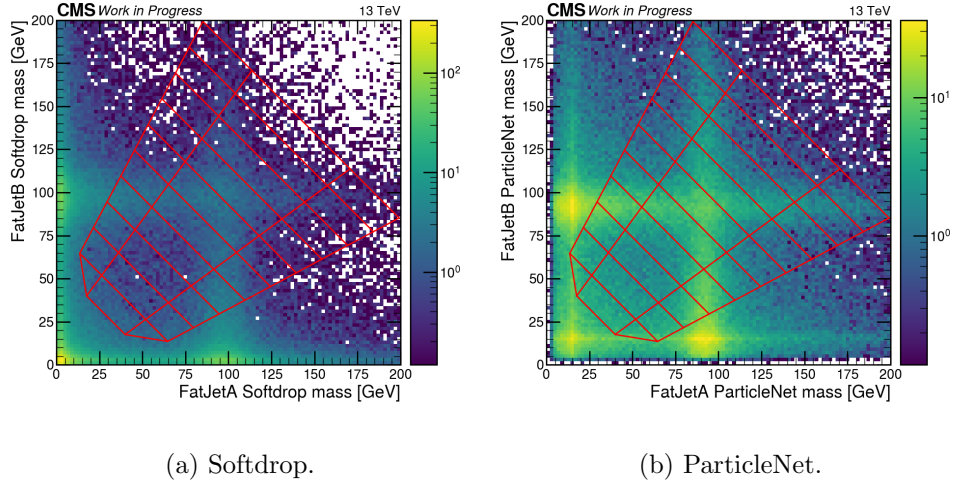


Figure 6.18: Distribution of the reconstructed jet mass for the two selected AK8 jets for the Z +jets background, for the Softdrop algorithm (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection were included.

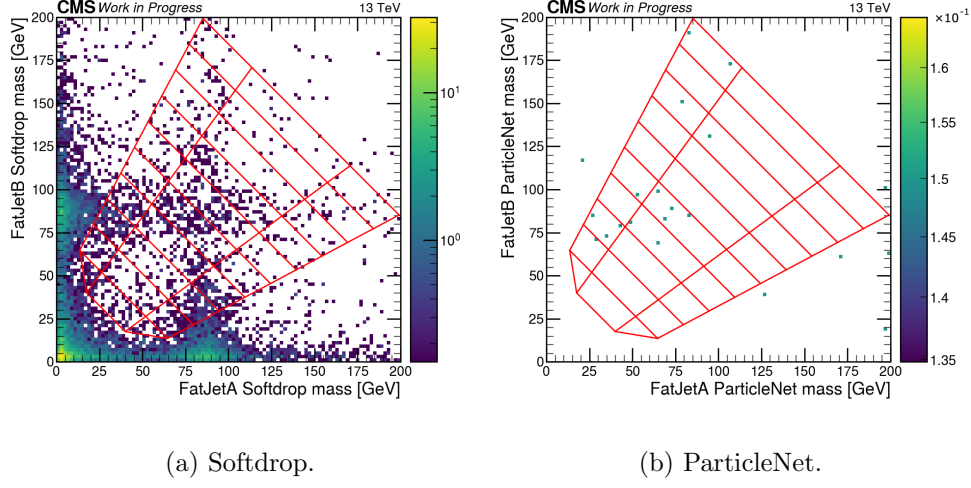


Figure 6.19: Distribution of the reconstructed jet mass for the two selected AK8 jets for the W +jets background, for the Softdrop algorithm with btagHbb (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection and the double- b -tag selection were included.

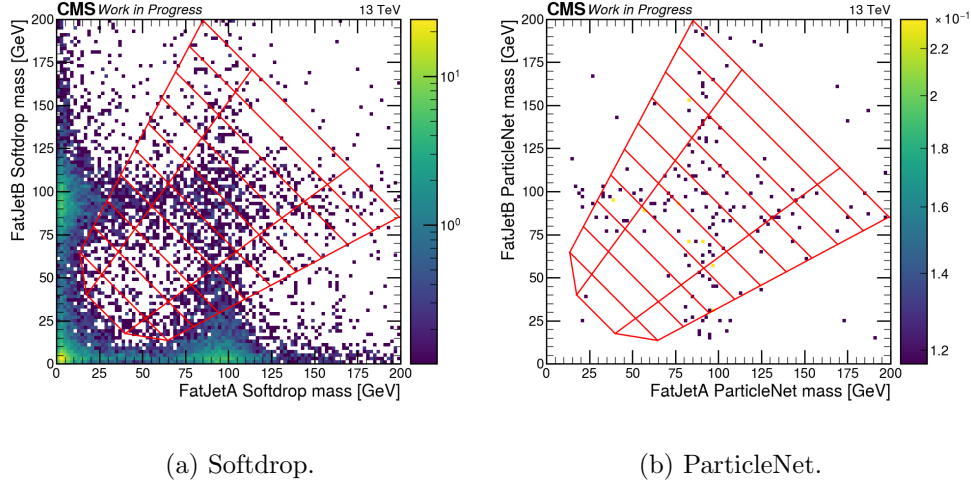


Figure 6.20: Distribution of the reconstructed jet mass for the two selected AK8 jets for the Z +jets background, for the Softdrop algorithm with btagHbb (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection and the double- b -tag selection were included.

6.7.4 Signal Mass Region

The yields in each signal mass region after the event selections, estimated using MC background and signal samples can be seen in Figure 6.21. It is evident that the number of background events found in the signal regions when using the ParticleNet algorithm is much lower than when using the btagHbb and Softdrop algorithms, while also keeping the potential signal yields high. In the event selection comparison between btagHbb and ParticleNet in the previous sections, an H_T cut of 1500 GeV was used to reproduce the results from the published analysis [12]. For the yield plots shown below, the selection requirement on the minimum H_T was increased to 2000 GeV because of a limited number of simulated events in the MC samples at low H_T bins which resulted in high statistical fluctuations. In particular, the improved double- b -tagging performance of ParticleNet dramatically reduced the number of $t\bar{t}$ events after the double- b -tag cut. For H_T above 2000 GeV the number of events in the samples were sufficient. All MC samples were centrally produced by CMS, however, samples containing more events need to be produced to extend the study to lower values of H_T .

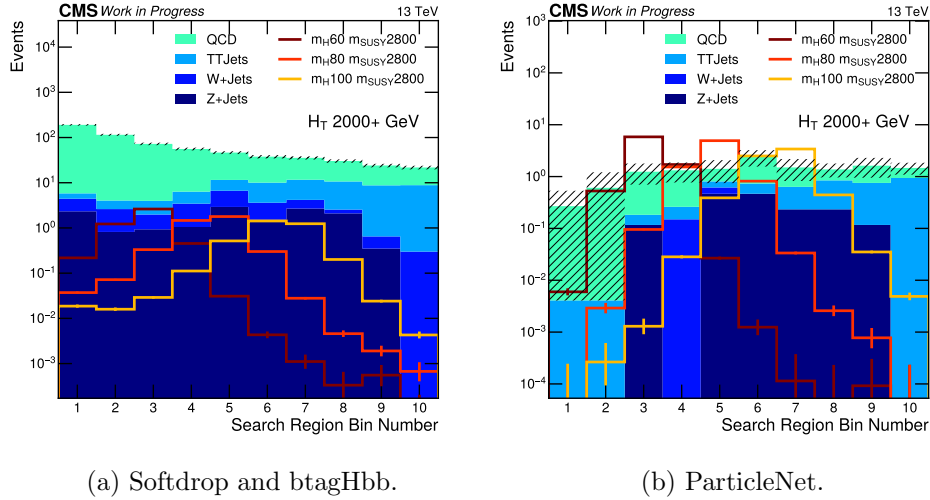


Figure 6.21: The expected yields in the TR for each signal mass region after the event selections for the btagHbb and Softdrop algorithms (left) and the ParticleNet algorithm (right). Both signal and background yields plotted here were estimated using simulated events. The hatched bands represent the statistical uncertainty on the total background. No systematic uncertainties are shown.

6.8 Background Estimations

For backgrounds with small contributions, such as the W +jets and Z +jets backgrounds in this study, the background count in each signal bin is estimated using only the MC samples. Although such predictions from MC have large relative systematic uncertainties, their contribution to the total background uncertainty is small. For larger backgrounds, it is more important to have a precise estimate, and an approach based as much as possible on measurements in data (data-driven) is often desirable. This can eliminate many systematic uncertainties inherent in a MC-based estimate, both experimental such as detector efficiency and theoretical systematics such as cross section. In the previous analysis, the contribution of the $t\bar{t}$ background was very small compared to that of QCD, so it was sufficient to estimate it from simulation. However, since ParticleNet is much better at rejecting QCD background, this may no longer be the case. A potential data-driven background estimation for the $t\bar{t}$ background is discussed in Section 6.11.3.

6.8.1 QCD Data-driven Background Estimation

The QCD background, which is the largest background, is estimated using data and the sideband regions U_i defined in Figure 6.10. The method is an “ABCD” method [88]. Each signal region S_i is designed to be wide enough that only a negligible number of signal events should end up in the U_i sideband region. The size of the sideband regions was chosen to have the same area as its corresponding signal region, thus if the density of events is approximately uniform along the sideband U_i and the signal region S_i , then the number of QCD background events in the two regions should be approximately equal. However, it is not a requirement of the ABCD method for the event counts to be equal, as long as the distribution of the density of events is similar before and after the double- b -tag cut (Figures 6.13-6.14). The event yields in the signal and sideband regions are denoted $\hat{S}_i$ and $\hat{U}_i$, respectively. To make sure only the background events are counted and do not include any signal, only events that fall into the CR are used. The ratio (F_i) between the event yields of S_i and U_i in the CR is defined as

$$F_i := \frac{\hat{S}_i^{\text{CR}}}{\hat{U}_i^{\text{CR}}}. \quad (6.3)$$

The QCD background in the TR can then be estimated as

$$\hat{S}_i^{\text{TR}} = F_i \hat{U}_i^{\text{TR}}, \quad (6.4)$$

where $\hat{U}_i^{\text{TR}}$ is the observed yield in the TR after subtracting the MC simulated background estimations (currently from $t\bar{t}$, Z +jets, and W +jets). This background estimation is only valid if there is no, or minimal, correlation between

the double- b -tag discriminator and the mass, such that the ratio S_i/U_i is the same in the CR and the TR. In the previous analysis, this was found to be the case when using btagHbb and the Softdrop mass. The ParticleNet double- b -tagger is explicitly trained to be uncorrelated with the jet mass, by using a flat distribution of Higgs masses in the training. However, some residual correlation may be present, so the background estimation was verified using a similar method as was used for btagHbb in the previous analysis: by comparing the predicted $\hat{S}_i^{\text{TR}}$ and $\hat{S}_i^{\text{VR}}$ to the simulated yield in the QCD MC sample, and by comparing the simulated and predicted yields in the VR using data. The results from this study are presented below.

Estimation from QCD MC

When using solely simulated QCD background events for validating the background estimation, no background events from other processes need to be considered when predicting the yields. A ParticleNet comparison of the calculated yield $\hat{S}_i^{\text{TR}}$ and the observed TR yield in the simulated QCD sample can be found in Figure 6.22 for the default double- b -tag sum cut of 1.95 and a loose cut of 1.9. The same QCD MC samples were used as for the event selection studies in Section 6.7. Because of the good performance of the ParticleNet double- b -tag discriminator, the yields in each bin are very small, which in turn results in large fluctuations and statistical errors for the estimations, even for the looser cut. Although the predictions from the method are somewhat consistent with the true QCD yields, the large uncertainties do not allow us to probe the self-consistency of the method with great conviction.

For estimating the yield $\hat{S}_i^{\text{VR}}$ using the QCD MC, Equation 6.4 was modified into the following equation:

$$\hat{S}_i^{\text{VR}} = F_i \hat{U}_i^{\text{VR}}, \quad (6.5)$$

where $\hat{U}_i^{\text{VR}}$ is the observed yield. The F_i factors were still calculated from the CR, as in Equation 6.3, but using only the simulated QCD sample. The QCD background estimation in the VR can be seen in Figure 6.23. Because the VR region contains more events, there was no issue with large fluctuations or statistical errors and the estimation seems to be working well, even for larger H_T cuts.

Estimation from Data

For the validation using data, $\hat{U}_i^{\text{VR}}$ in Equation 6.5 now represents the yield in data after subtracting the yields from the MC simulated backgrounds (excluding QCD). As no data-driven $t\bar{t}$ background estimation has been implemented, an estimation of it was made using the MC background. This is justified because the proportion of $t\bar{t}$ events in the VR is much smaller than in the TR. The estimated yields for the default double- b -tag discriminator cut can be seen

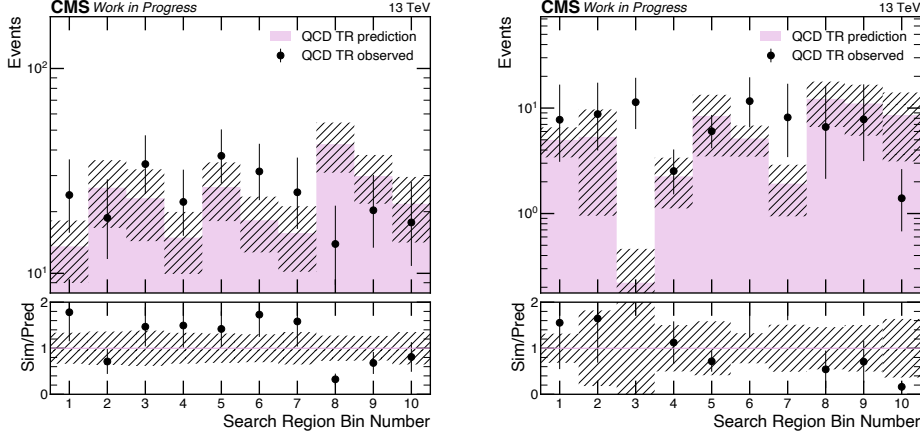
(a) Loose double- b -tag cut (1.90).(b) Default double- b -tag cut (1.95).

Figure 6.22: A comparison of the predicted yields and the observed yields in the TR of the QCD MC background for a loose double- b -tag cut (left) and the default double- b -tag cut (right). The hatched bands and error bars represent the statistical uncertainty for the background prediction and observed yields, respectively. No systematic uncertainties are shown.

in Figure 6.24 and show that the QCD background estimation is performing well. The results are also comparable to what was observed in the published analysis [12], where a systematic uncertainty for the VR closure test of 15% was assigned to H_T categories < 3500 GeV, and a 30% systematic uncertainty to $H_T \geq 3500$ GeV [12]. Studies were performed for the published analysis using btagHbb to ensure that any differences between the background estimation in the VR and the TR were negligible. A similar study needs to be performed for ParticleNet.

6.9 Systematic Uncertainties

The analysis is affected by various systematic uncertainties, arising from the use of simulated signal and background events and jet reconstruction, for example. These uncertainties must be studied to determine the effect on the final outcome of the analysis. If the uncertainty is too large, the results will be inconclusive. The total uncertainty in the published analysis [12] was dominated by the statistical uncertainty from the data TR yields in the S_i and U_i regions, and the systematic uncertainties with the largest contributions were attributed to the following sources:

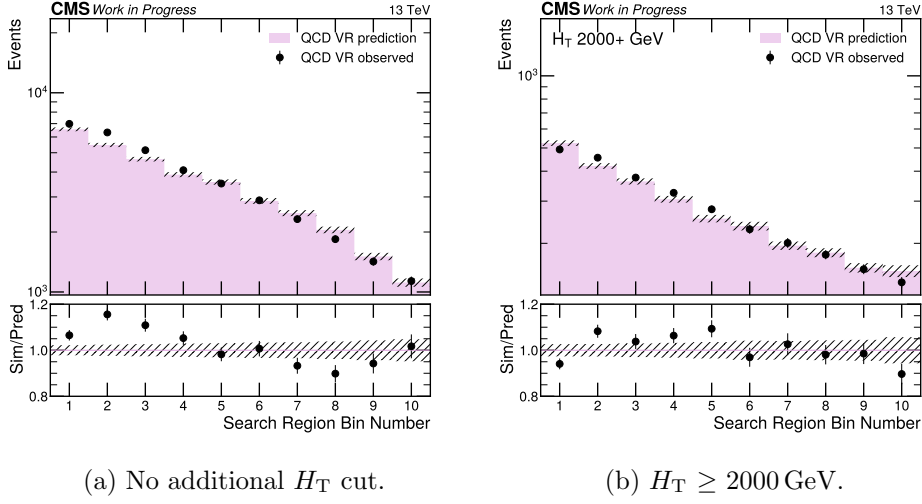


Figure 6.23: A comparison of the predicted yields and the observed yields in the VR of the QCD MC background, using the default double- b -tag discriminator sum cut (1.95). The estimations were made using no additional H_T cut beyond that applied by the trigger (left) and using $H_T \geq 2000$ GeV (right). The hatched bands and error bars represent the statistical uncertainty for the background prediction and observed yields, respectively. No systematic uncertainties are shown.

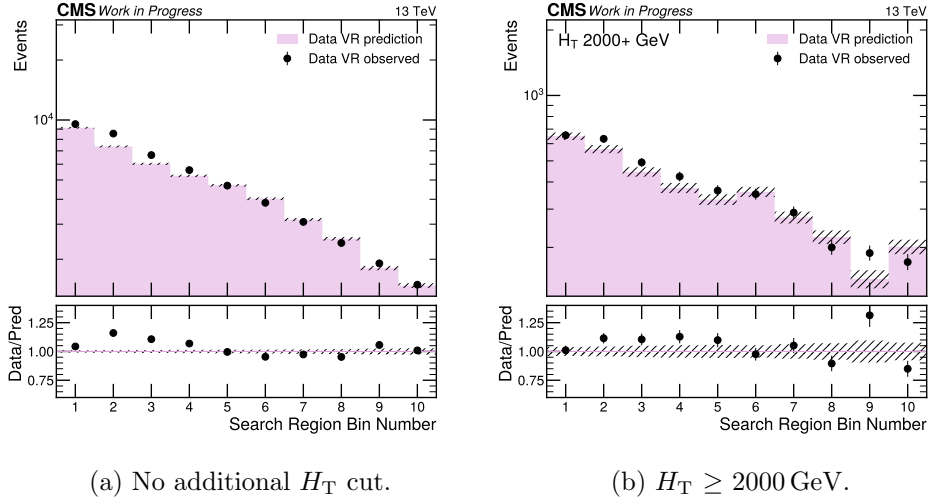


Figure 6.24: A comparison of the predicted yields and the observed yields in Run 2 data from 2018 in the VR, using the default double- b -tag discriminator sum cut (1.95). The estimations were made using no additional H_T cut beyond that applied by the trigger (left) and using $H_T \geq 2000$ GeV (right). The hatched bands and error bars represent the statistical uncertainty for the background prediction and observed yields, respectively. No systematic uncertainties are shown.

- Background estimations
- Double- b -tag efficiency
- Jet Mass Scale (JMS) and Jet Mass Resolution (JMR)
- Jet Energy Scale (JES) and Jet Energy Resolution (JER)

The data-driven QCD background estimation gives rise to systematic uncertainties on the measured F_i factors (Equation 6.3) which have currently not been estimated for this analysis. Figures 6.22-6.24 suggest that these uncertainties are likely to be similar to the published analysis, i.e. 15-30%. However, this needs to be assessed. Additionally, if a data-driven background estimation is implemented for $t\bar{t}$, its systematic uncertainty must be assessed too.

The double- b -tag efficiency, JMS and JMR, all contribute to uncertainties related to ParticleNet in the improved analysis. The systematic uncertainties which are directly affected by ParticleNet are discussed in Sections 6.9.1-6.9.2. These uncertainties had to be implemented manually as NANOAODTOOLS did not include uncertainties related to ParticleNet. The JES and JER are used to account for discrepancies between the reconstructed AK4/AK8 jet energy and the jet energy from the true value generated in the MC, as well as discrepancies between the reconstructed jets in MC and data. The built-in NANOAODTOOLS JES and JER systematic uncertainties using values recommended by CMS were employed for this purpose.

Additional systematic uncertainties that were found to be negligible in the previous analysis need to be reexamined to ensure they still do not have a significant impact. While pileup scale factors have already been included in NANOAODTOOLS, other sources of uncertainties will need to be added.

6.9.1 Double- b -tag Scale Factor

The purpose of the double- b -tag scale factor is to adjust for any discrepancies in the efficiency of tagging a jet that originates from two b quarks between simulated signal events and real data. The correction values and associated uncertainties are sourced from an analysis which used ParticleNet to identify jets from $HH \rightarrow 4b$ decays [89, 90], similar to the signal for this analysis. In that analysis, $b\bar{b}$ jets originating from gluons were studied in both a data control sample and MC. By comparing the efficiencies of tagging the $b\bar{b}$ pairs, a correction factor to better match the MC with data could be determined. This was done for several bins of double- b -tag discriminator values. To exclude any $b\bar{b}$ jets originating from gluons in the QCD background, a BDT was trained to identify and remove such jets. The calculated scale factor applies to a single double- b -tagged jet and depends on the double- b -tag discriminator value and the p_T . The total scale factor for an event is calculated by multiplying the scale factors of two AK8 jets that passed the kinematic selection. For each fat jet

in a simulated event that is matched to a generator-level $H \rightarrow b\bar{b}$ jet, the scale factor is applied to the event weight. The effect of the double- b -tag uncertainty is studied by scaling the event weight by the scale factor and its corresponding uncertainty, which results in $7.4^{+5}_{-3.7}$ signal events that pass the final selection for $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV. The large uncertainties from the ParticleNet double- b -tag scale factors are expected to reduce in the future as the methods of measuring the scale factors for this novel double- b -tagger become more mature. As each chosen AK8 jet is treated independently, no correlation between the two scale factors is assumed.

6.9.2 Jet Mass Systematic Uncertainties

The systematic uncertainties associated with the ParticleNet jet mass involve the application of scale factors and mass resolution corrections to the selected AK8 jets' masses. The JMS correct the mean of the reconstructed jet mass distribution in the MC events to match the mass distribution in data, while the JMR adjusts the spread of the reconstructed jet mass distribution. Both JMS and JMR values affect the reconstructed jet mass (m_{reco}), thus they are applied simultaneously to obtain the corrected jet mass ($m_{\text{corrected}}$) using the equation:

$$m_{\text{corrected}} = m_{\text{reco}} \cdot \text{JMS} - \Delta m \cdot \text{JMR}, \quad (6.6)$$

where Δm is the mass difference between the generated mass and m_{reco} . The JMS and JMR values used in the analysis were taken from a di-Higgs analysis with a boosted $4b$ final state [91, 92]. That analysis used the ParticleNet mass regression to reconstruct the AK8 jet masses from the $b\bar{b}$ pairs, and the JMS and JMR values were obtained from a fit of the target mass distribution. The ParticleNet jet mass corrections were only applied to MC signal events.

To study the overall effect of the jet mass systematic uncertainties, the associated uncertainties for JMS and JMR were applied separately to Equation 6.6. In other words, only either the JMS or the JMR was shifted up or down to evaluate the corresponding uncertainty. The JMS and the JMR mainly shift the signal events between signal mass bins, but varying the JMR also shifts events between S_i and U_i . The effects of JMS and JMR on the mass distribution can be seen in Figure 6.25. For a signal sample with $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV, the migration between adjacent signal mass regions were $< 1.5\%$ when varying JMS and $< 4.9\%$ when varying JMR. The migration to the sideband regions was small ($< 0.1\%$) for both systematic uncertainties, due to the good mass resolution of ParticleNet yielding fewer events near the signal and sideband region border. The effects were greater for a signal with $m_{H_1} = 30$ GeV and $m_{\text{SUSY}} = 2800$ GeV, with a migration of $< 4.7\%$ for JMS and $< 5.2\%$ for JMR, where almost all events migrated outside the mass grid. The corresponding values for small m_{H_1} in the published analysis were up to 10% for JMS, and up to 20% for JER [12]. The fewer migrations for ParticleNet can be explained by the overall better resolution

for ParticleNet and an improved JMR variation method that makes use of the generated mass. As in any analysis based on limit settings of model parameters, it's important to check correlation assumptions between uncertainties.

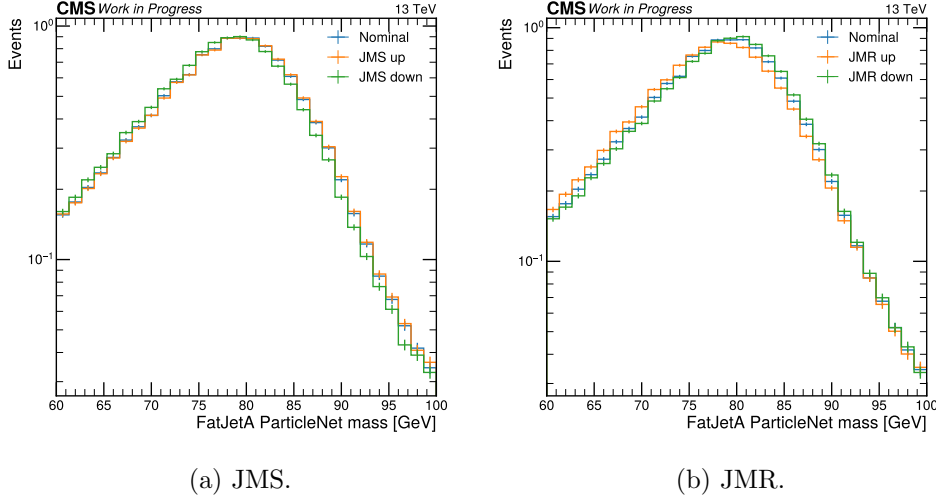


Figure 6.25: The effect of the JMS (left) and JMR (right) uncertainties on the nominal corrected mass for a signal sample with $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV.

6.10 ROC Curves

For an easier comparison between the two double- b -taggers, Receiver Operating Characteristic (ROC) curves of the double- b -tag performances have been plotted in Figure 6.26, using the event counts before the mass event selection. The ROC curves show the efficiency of correctly tagged signal jets (true positives), compared to the rate of mistagged background jets (false positives), over the range of possible double- b -tag discriminator sum cuts. It is desired to choose a double- b -tag cut that has a high rate of correctly tagged jets compared to mistagged jets. The ParticleNet tagger shows a better ROC performance than the btagHbb tagger, measured against the QCD, $t\bar{t}$, W +jets, and Z +jets backgrounds. It can also be observed that fat jets in $t\bar{t}$ events are more likely to be mistagged. However, as QCD is still the dominant background in most of the phase space of the ROC curve, the ROC curve for the combined backgrounds follows the QCD line. The results do not depend significantly on the choice of signal sample. The kink for ParticleNet which is mainly visible for the $t\bar{t}$ background corresponds to a double- b -tag sum cut of 1 and is due to the shape of the distribution of the double- b -tag sum which has a minor peak

at 1 (i.e. when in general one fat jet was assigned a double- b -tag discriminator value close to 0 and the other fat jet was assigned a value close to 1).

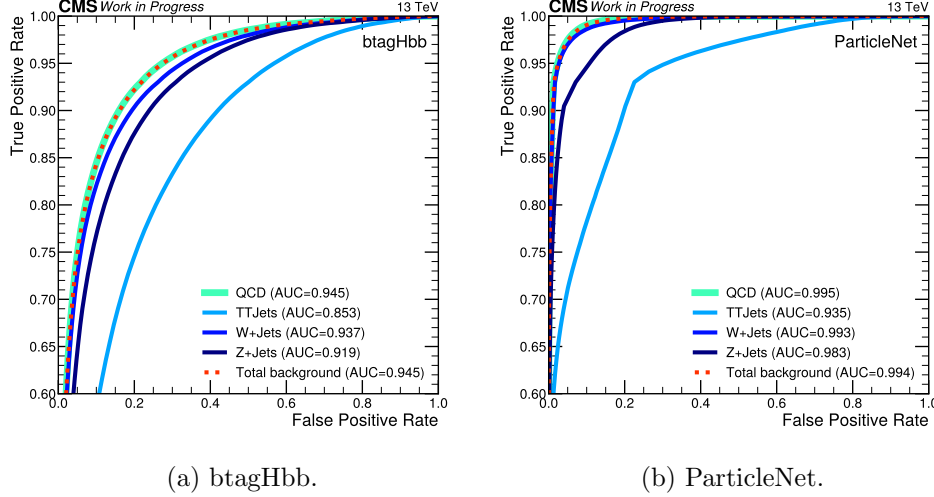


Figure 6.26: The ROC curve for a range of double- b -tag discriminator sum cuts for the btagHbb (left) and the ParticleNet (right) algorithms. These values were obtained using the QCD, $t\bar{t}$, W +jets, and Z +jets background samples, and a signal sample with $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV. The Area Under the Curve (AUC) is a general measure of the discriminator performance over all possible classification thresholds.

The improved double- b -tagging performance using ParticleNet can also be seen in Figure 6.27, where the number of mistagged jets is several magnitudes less in ParticleNet than for btagHbb. It can also be seen that the ratio of the number of QCD to $t\bar{t}$ events has been reduced, going from a factor ~ 100 to ~ 10 . This improvement is observed when only the double- b -tag selection is considered. Also note that before the mass event selection, the W +jets and Z +jets yield a similar number of false positives as the $t\bar{t}$ background for btagHbb.

Using the ROC curves, an attempt was made to choose the best double- b -tag cut for the ParticleNet discriminator. The first attempt involved maximising the discovery significance using a simple equation that is valid for when $s \ll b$

$$Z = \frac{s}{\sqrt{b}}, \quad (6.7)$$

where Z is the significance, s is the number of signal events, and b is the number of background events. However, due to the good performance of the ParticleNet discriminator, a substantial fraction of signal events still remain at the point at which no MC background events pass the cut (Figure 6.27). Therefore, Equation 6.7 was always maximised when no background passed

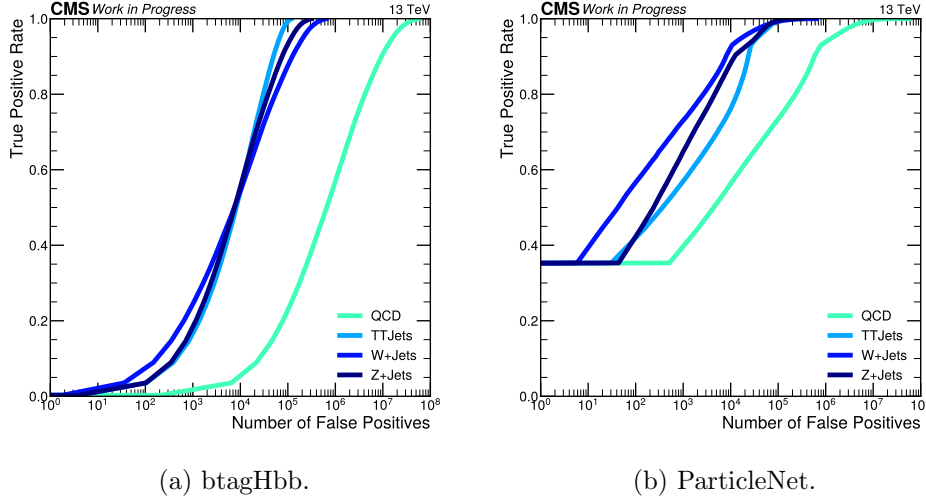


Figure 6.27: The number of mistagged jets for a range of double- b -tag discriminator sum cuts for the btagHbb (left) and the ParticleNet (right) algorithms. These values were obtained using the QCD, $t\bar{t}$, W +jets, and Z +jets background samples, and a signal sample with $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV. The horizontal lines at a low number of false positives happen when the true positive rate is greater than zero, but there are no false positives (i.e. no MC background events pass the cut).

the cut. Due to the lack of MC data in this region, such a cut was considered an unreasonable choice. An attempt at optimising the double- b -tag cut was also made using a more sophisticated equation compared to Equation 6.7, for exclusion significances [93]:

$$Z = \sqrt{2((s+b)\log(1+s/b) - s)}. \quad (6.8)$$

Unfortunately, the more sophisticated figure of merit was similarly affected as Equation 6.7. The ROC curves in Figure 6.26 were created using event counts prior to the mass event selection. However, this approach may be suboptimal for determining the appropriate double- b -tag cuts, as the primary concern lies in evaluating the signal and background yields after all selection criteria are applied. An attempt to determine the optimal cut on the sum of the discriminators was also made from ROC curves created after the mass event selection. In Figure 6.28, it can be seen that the ParticleNet performance is superior and that the W +jets and Z +jets have less of an impact on the false positive rate after the mass event selection. Again, it was not possible to precisely determine the optimal cut using such a study. It is worth noting that the number of QCD and $t\bar{t}$ events when using ParticleNet only become similar for large double- b -tag cuts, i.e. the left side of the ROC curve. A more realistic study of optimal cuts is left for future work when all analysis cuts can

be simultaneously optimised for discovery. This is discussed further in Section 6.11.1.

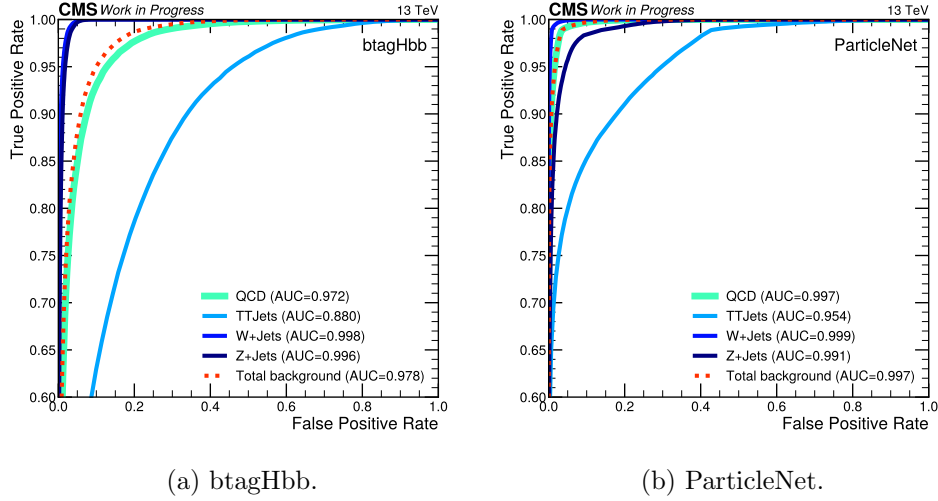


Figure 6.28: The ROC curve for a range of double- b -tag discriminator sum cuts for the btagHbb (left) and the ParticleNet (right) algorithms. These values were obtained after all event selections using the QCD, $t\bar{t}$, W +jets, and Z +jets background samples, and a signal sample with $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV. The Area Under the Curve (AUC) is a general measure of the discriminator performance over all possible classification thresholds.

6.11 Improvements

The adoption of ParticleNet, as discussed in this chapter, has led to significant improvements in the analysis. However, there remains potential for future refinements. As the analysis is still ongoing and data-taking for Run 3 is in progress, final limits on the cross section and branching fraction of H_1 have not yet been established. Limit calculations will be conducted in the future using the Combine Tool [94], however, it is important to study additional systematic uncertainties beyond those implemented in Section 6.9 to perform the complete analysis. Further improvements and future work are outlined in this section, highlighting areas for potential enhancements and advancements.

6.11.1 Double- b -tag cuts

With the change of the double- b -tag discriminator, the double- b -tag regions in Figure 6.3 need to be reoptimised for ParticleNet. The most important change involves the TR definition, but the size and shape of the CR and the VR can also be improved. The TR needs to be defined such that the background events

in the final selection are few, while the number of signal events remains high. The initial attempt to determine the optimal TR using ROC curves in Section 6.10 was unsuccessful, thus a more comprehensive study is needed. The new study would involve conducting the complete analysis using MC samples and optimising all cuts (particularly, for the double- b -tag and mass event selections) simultaneously for discovery significance. The CR and VR both need to have a high QCD purity with negligible signal contamination for the QCD data-driven background estimation. The fraction of other background events, such as $t\bar{t}$, also needs to be small such that it can be subtracted from the background estimation with an uncertainty that is small compared to the overall uncertainty of the ABCD method. All of these criteria for the double- b -tag regions were met in the published analysis. Although ParticleNet is showing great improvements compared to btagHbb, the two algorithms' double- b -tag distributions show similar features, thus similar double- b -tag regions can be used for both double- b -taggers. By executing the full analysis and incorporating all systematic uncertainties with e.g. varying TRs, it is possible to assess which cuts will yield the most favourable outcomes in terms of final results.

6.11.2 Mass Region Bins

Because the mass resolution of the ParticleNet mass is better than that of the original Softdrop mass, the signal mass regions can be split into two bins: one central region on the diagonal and one region combining the area on either side of the central signal region. Since the signal distributions are approximately Gaussian in the 2D mass plane, signal events are concentrated towards the diagonal. Splitting the signal region into two bins will yield one bin with a higher signal-to-background ratio, thus making any signal excess in data more prominent there. This adjustment also allows for shifting the sideband regions further away from any signal events which would interfere with the data-driven background estimation. A visual representation of an updated mass region grid can be seen in Figure 6.29.

6.11.3 Data-driven Background Estimation for $t\bar{t}$ Background

Because of the significant improvement in QCD background rejection achieved when using ParticleNet, the contributions from QCD are reduced to a level where $t\bar{t}$ is also significant. The $t\bar{t}$ background is now only about a factor of 10 smaller than QCD. Consequently, a data-driven background estimation method is desirable for the $t\bar{t}$ background as well. The ABCD method used for QCD background, which relies on the F_i factors (Equation 6.3) being the same in the CR and TR as a consequence of the mass and double- b -tagger being uncorrelated, can no longer be used as this is not the case for $t\bar{t}$. An alternative method must, therefore, be used for a data-driven $t\bar{t}$ background estimation.

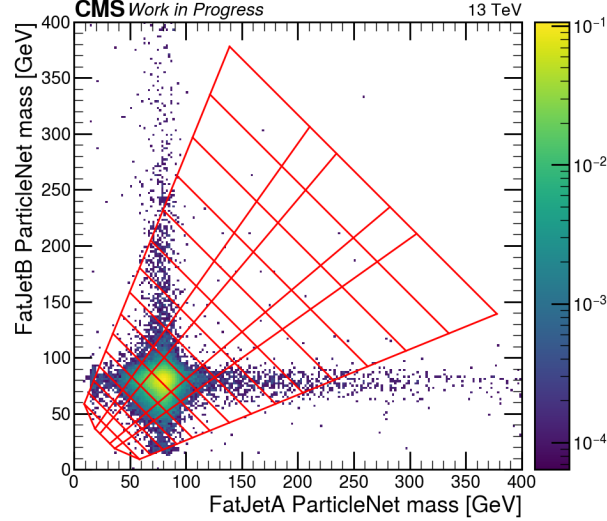


Figure 6.29: An example of mass region bins for ParticleNet with two signal regions: the central bins where most signal events will occur, and the surrounding bins. The outermost bins are the sideband regions for the data-driven background estimation. The plot was made using signal with $m_{H_1} = 80$ GeV and $m_{\text{SUSY}} = 2800$ GeV after both the kinematic and double- b -tag event selection.

A potential approach to perform a data-driven background estimation for the $t\bar{t}$ background involves adapting the data-driven method by introducing separate CR and VR regions specifically designed for $t\bar{t}$. For example, the regions can be defined in the 2D space of the ParticleNet top-tag discriminator. Similar to the ParticleNet double- b -tagger, the top-tagger uses a neural network to classify AK8 jets, but this time from a hadronic top quark decay. By constructing appropriate criteria for the CR and VR, tailored to capture the characteristics of the $t\bar{t}$ background, it should be possible to use an alternative data-driven method to accurately estimate the $t\bar{t}$ background contribution.

6.12 Conclusions

This study explores how a new double- b -tagger, ParticleNet, would perform in an NMSSM physics search. The NMSSM analysis focuses on a search for two boosted light Higgs bosons decaying into b -quark pairs in all-hadronic events with little missing transverse momentum. The goal is to improve upon a published CMS analysis from Run 2 and adapt it for the Run 3 data. To enhance the sensitivity of the analysis, ParticleNet was employed to improve

the double- b -tagging performance. The use of ParticleNet not only improved the background rejection capabilities but also allowed for a more efficient selection of signal events. The greatest improvements were observed for the QCD background, the main background for the signal model, which could be reduced by a factor of 100 compared to the double- b -tagger used in the published analysis (btagHbb), if only the double- b -tag discriminator was considered. Additionally, the ParticleNet mass was introduced for jet mass reconstruction, improving mass resolution and largely eliminating signal jets previously mistaken as massless. This resulted in twice as many signal events passing the selections. Using ParticleNet throughout the analysis leads to fewer background events per search region bin while maintaining the desired signal yield. These remarkable advancements lead to an increased sensitivity of the analysis, enabling the study of larger m_{SUSY} parameter values beyond those examined in the published analysis. Furthermore, optimising the signal event selection criteria to align with the capabilities of the new ParticleNet double- b -tagger and jet mass reconstruction will further improve the overall analysis sensitivity.

To conduct these ParticleNet performance comparisons, 2018 MC simulations were used. An attempt at optimising the cuts on the new double- b -tagger for the improved analysis was made, however, this was foiled by a lack of simulated QCD events and more advanced techniques need to be used. The MC simulations and 2018 data were also used to validate the data-driven QCD background estimation based on the approach used in the previous analysis. The studies confirmed the viability of using ParticleNet for the QCD data-driven background estimation and demonstrated the successful implementation of the estimation.

While an improved background rejection when using ParticleNet was observed for all sources of background, the significant enhancement in QCD background rejection made the $t\bar{t}$ background more prominent in comparison to the published analysis. The contributions from $t\bar{t}$ are no longer negligible compared to QCD, thus a dedicated data-driven method can be developed specifically for the estimation of the $t\bar{t}$ background to enhance the accuracy of the background estimation.

To finalise the analysis framework, additional systematic uncertainties should be added and studied for the Run 3 data. Some systematic uncertainties for the analysis have been implemented, such as those from the double- b -tag scale factors and jet mass uncertainties. However, further investigation is necessary to comprehensively explore these uncertainties. By addressing the systematic uncertainties and refining the analysis techniques as mentioned earlier, the analysis performance can be maximised. ParticleNet has already shown significant improvements compared to the published analysis, which will enable more precise and reliable results from the Run 3 data, advancing the search for any potential light Higgs boson decays.

Chapter 7

Conclusions

The field of particle physics remains captivating and continuously evolving. Even though all particles predicted by the Standard Model have been discovered, there is still much to learn by continuing precise measurements of the model, as well as searching for particles beyond the Standard Model. This thesis describes various particle physics research initiatives within the CMS experiment at the Large Hadron Collider (LHC). These efforts encompass a wide range of projects: from preparing CMS for the High-Luminosity LHC (HL-LHC) upgrade, to improving a study of the Next-to-Minimal Supersymmetric Standard Model (NMSSM) in the hunt for undiscovered light Higgs bosons.

The HL-LHC will allow a high data-collection rate by increasing the number of simultaneous proton-proton collisions up to 200. At CMS, the hardware-based L1 Trigger (L1T) system is the first step in deciding which events to retain and which events to discard. The rate at which the L1T can accept an event is driven by the constraints of data transmission and storage capacity. However, the current L1T would not be able to cope with the increased rate at the HL-LHC. Most proton-proton collisions are uninteresting so-called pileup events, thus a more efficient L1T is needed to identify the interesting events for physics analysis. The pioneering work on the L1 tracking-finder system for the HL-LHC is expected to significantly reduce the L1T rate while enhancing the detector's sensitivity to physics at higher luminosities. For CMS, it marks the first time that data from one of the subdetectors closest to the proton collisions, the Outer Tracker, will be used in the L1T. The L1 Track Finder aims to reconstruct all charged particles with $p_T > 2$ GeV in under 4 μ s thanks to an FPGA-implemented algorithm. This algorithm consists of various processing modules implemented in firmware, which are connected to form a track-finding chain. The individual processing modules have successfully met the stringent timing requirements and passed firmware simulations. Subsequent tests involving a slice of the full track-finding chain have also been successfully demonstrated. Work is ongoing to expand the tests to include a

substantial portion of the track-finding chain. Moreover, the introduction of a modified algorithm variant that combines processing modules for further time and resource reduction shows promise for the project's future.

As mentioned, the L1 Track Finder can reconstruct tracks with a p_T down to 2 GeV. Increasing the minimum track p_T to 3 GeV would considerably reduce the number of tracks to reconstruct, which could save time and resources in both the L1 Track Finder and the L1T, at the cost of potentially degrading the L1T performance. A study of the L1T was made, and the differences between a 2 GeV and 3 GeV track p_T threshold were small for high- p_T triggers. It was also shown that including tracks in the L1T does indeed improve the trigger efficiencies.

The CMS back-end system plays a pivotal role in controlling and monitoring the detector. It also includes the data acquisition systems and the L1T. To handle the increased data rate at the HL-LHC, the back-end systems also need to be upgraded. One of these crucial upgrades involves the deployment of the “Serenity” back-end boards, which will be responsible for communicating with e.g. the Outer Tracker and for transmitting data to the L1 Track Finder. The data is transmitted via optical fibres, necessitating the use of optical transceivers to convert electrical signals into optical signals and vice versa. A study of prototype transceivers, called “Fireflies”, was performed which included extensive performance scans across different configurations of the Firefly prototypes. These 12-channel transceivers, which are designed to operate at 25 Gb/s, are essential in managing the substantial data volumes generated at the HL-LHC. The majority of the Firefly transceivers showed stable behaviour. Occasional issues, particularly with dysfunctional channels, were identified and brought up with the manufacturer. Optical oscilloscope measurements ensured compliance with product specifications. Attenuation measurements further validated the suitability of these components and their ability to handle weak signals. Despite challenges encountered with a small subset of prototype parts, the vast majority exhibited satisfactory performance. Collaboration with the manufacturer is ongoing to resolve issues detected in a subset of the channels. This study showed that these transceivers could be viable for data transmission at the HL-LHC, but continued monitoring and testing are required to address the observed issues.

The NMSSM is a proposed theory beyond the Standard Model which offers solutions to some of the unanswered questions in the Standard Model, for example, concerning the Higgs boson mass. The NMSSM predicts a whole new set of particles, some of which could be detectable at the LHC. A previously published NMSSM analysis focused on a search for two boosted light Higgs bosons decaying into b quark pairs with minimal missing transverse momentum. No new light Higgs bosons were found, however, novel analysis techniques could enhance the search's sensitivity for the upcoming Run 3 CMS data. This thesis discussed the possibility of using the recent ParticleNet double- b -tagger for identifying the jets caused by the b quark

pairs in the NMSSM signal. Incorporating ParticleNet leads to a substantial improvement in double- b -tagging performance, as well as an improvement in background rejection capabilities. In particular, the primary QCD multi-jet background was reduced significantly when compared to the previously used double- b -tagger. Moreover, the introduction of the ParticleNet mass regression algorithm improved the mass reconstruction of the particles from which the $b\bar{b}$ jets originated, i.e. the light Higgs bosons. This resulted in an increased number of signal events passing the signal event selections while keeping the number of background events low. These advancements, led by ParticleNet, enable greater sensitivity of the analysis. Further refinements to the selection criteria for signal events, optimised for the ParticleNet double- b -tagger and jet mass reconstruction capabilities, hold the potential for even more significant improvements.

Appendix A

Firefly Alpha Transceiver Measurements

Parameter scans similar to those described in Section 5.4 of Firefly 12-channel transceivers were originally conducted on Alpha prototypes. Table A.1 lists the third version of Alpha (Alpha3) components, which shows that the transmitter firmware is identical to that of the Beta components presented in Table 5.1. The Alpha3 and Beta receivers have differing firmware versions.

Table A.1: The Alpha3 Firefly transceivers that were delivered to Imperial College London in 2022.

Name	Type	Serial Number	Firmware Version
IC Apr2022	Tx Alpha3	UA211205PB	00120606
IC Apr2022	Rx Alpha3	UA202509YQ	00130A01
IC May2022-A	Tx Alpha3	UA211205RN	00120606
IC May2022-A	Rx Alpha3	UA200801FE	00130A01
IC May2022-B	Tx Alpha3	UA211205S9	00120606
IC May2022-B	Rx Alpha3	UA202509YJ	00130A01
IC May2022-C	Tx Alpha3	UA211205P4	00120606
IC May2022-C	Rx Alpha3	UA200801FS	00130A01
IC May2022-D	Tx Alpha3	UA2109036M	00120606
IC May2022-D	Rx Alpha3	UA202509YT	00130A01

During the transceiver parameter scans conducted on the Alpha3 components, the receiver termination voltage was set to 800 mV. The scans were performed on two Serenity boards for comparison: Serenity 04 and Serenity HNZN. However, as shown in Figures A.1-A.5, performances are overall poor and inconsistent, with fewer than twelve links yielding no errors in the majority of all parameter combinations.

APPENDIX A. FIREFLY ALPHA TRANSCEIVER MEASUREMENTS

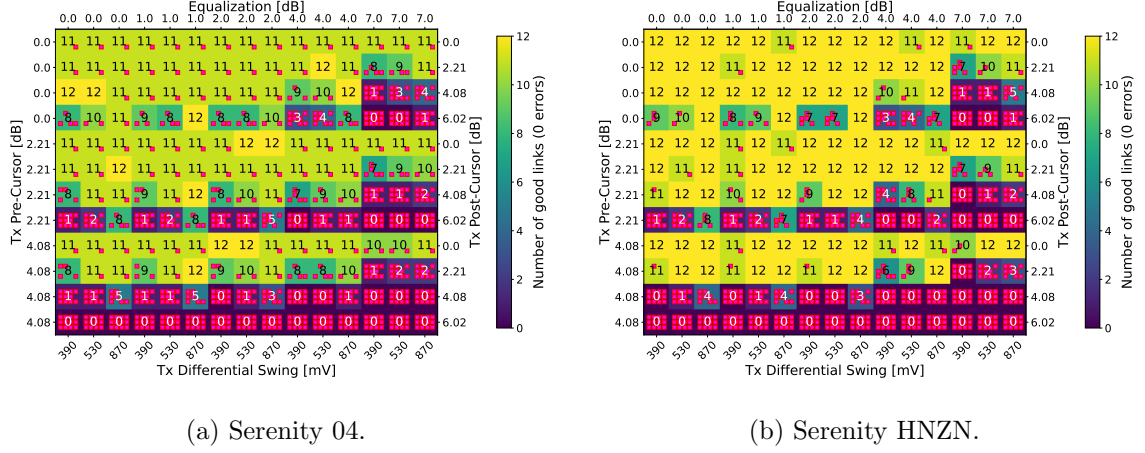


Figure A.1: Transmitter parameter scan results for ICapr22 on Serenity 04 (left) and HNZN (right).

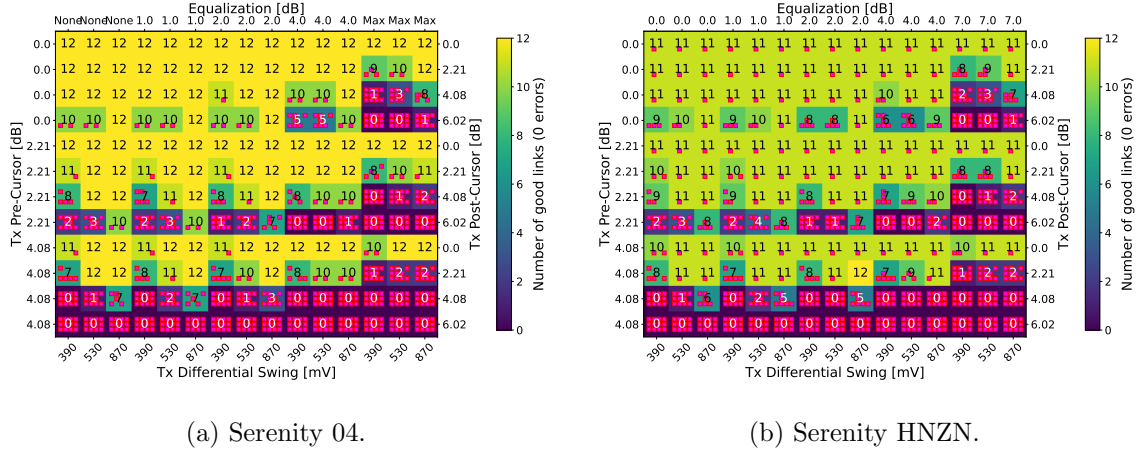


Figure A.2: Transmitter parameter scan results for ICmay22a on Serenity 04 (left) and HNZN (right).

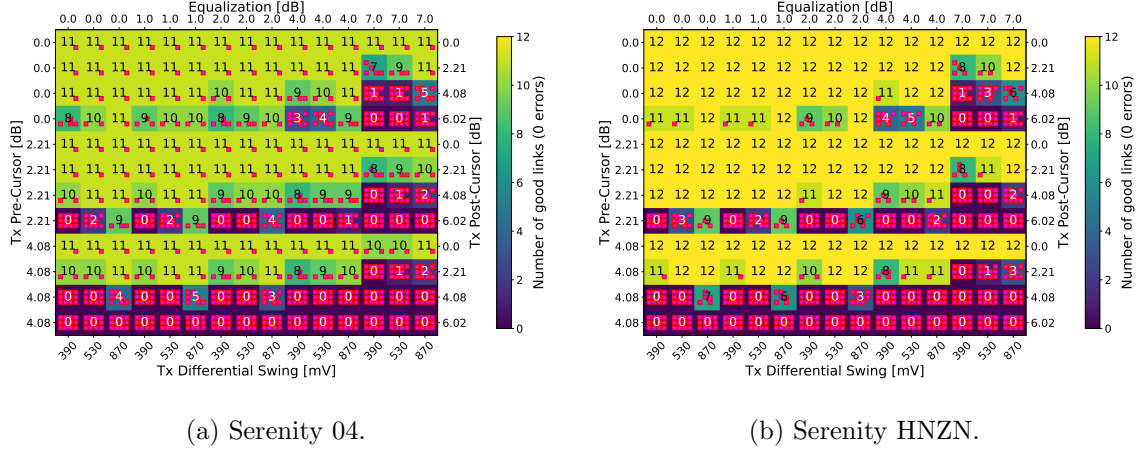


Figure A.3: Transmitter parameter scan results for ICmay22b on Serenity 04 (left) and HNZN (right).

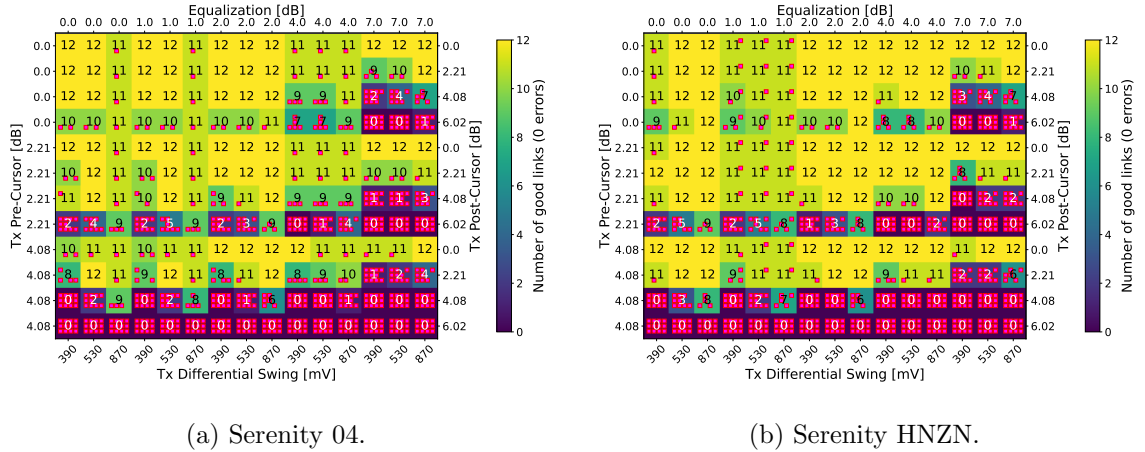
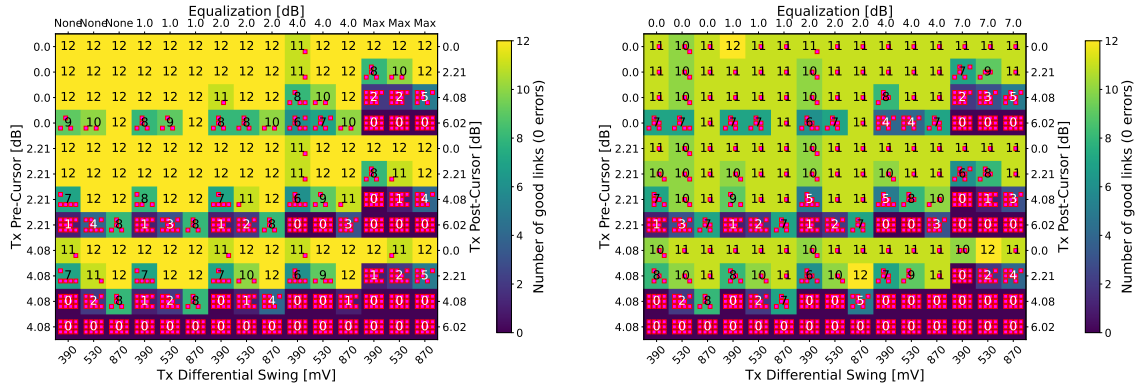


Figure A.4: Transmitter parameter scan results for ICmay22c on Serenity 04 (left) and HNZN (right).



(a) Serenity 04.

(b) Serenity HNZN.

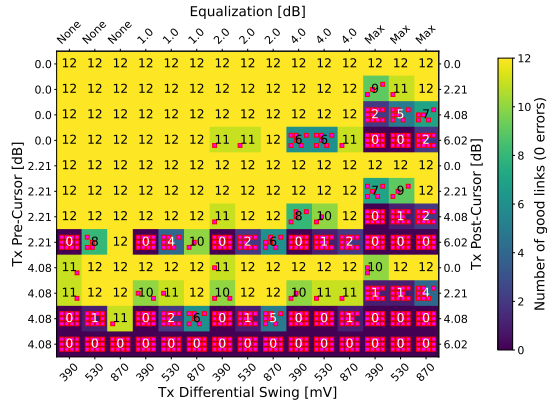
Figure A.5: Transmitter parameter scan results for ICmay22d on Serenity 04 (left) and HNZN (right).

Appendix B

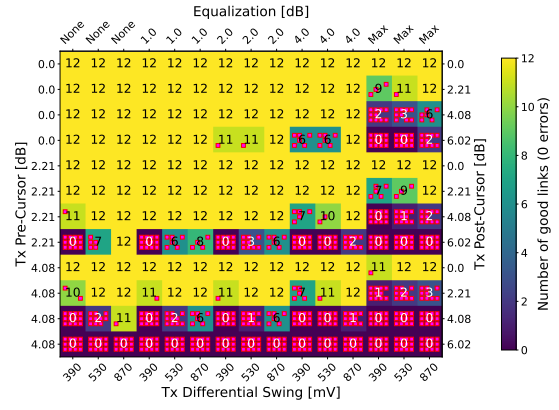
Firefly Beta Transceiver Measurements on Serenity HNZN

B.1 Parameter Scan

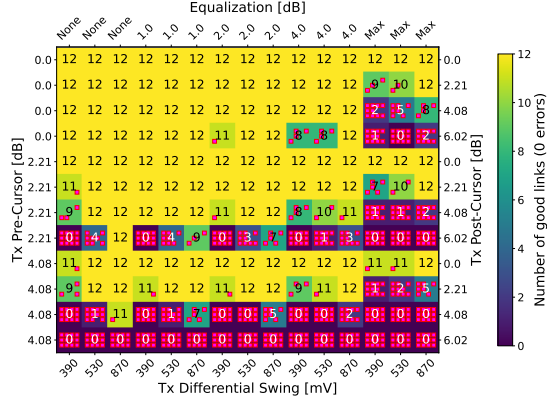
To make sure the parameter scan results from the 12-channel Beta parts (Section 5.4) were not dependent on the Serenity board used, the same scans were carried out on a second board named Serenity HNZN. The transceiver scans and bathtub curves obtained from the alternative Serenity board are shown in Figure B.1 and B.2, respectively. These outcomes align with the measurements taken on Serenity HNZN, including the CDR issue discussed in Appendix B.1.1.



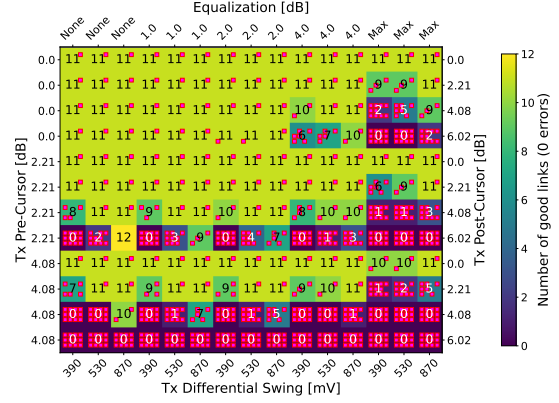
(a) ICjul22a.



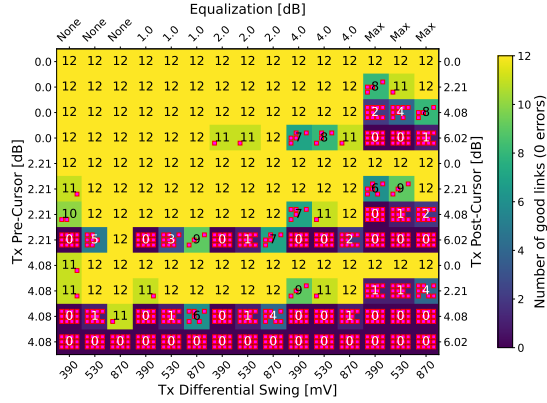
(b) ICjul22b.



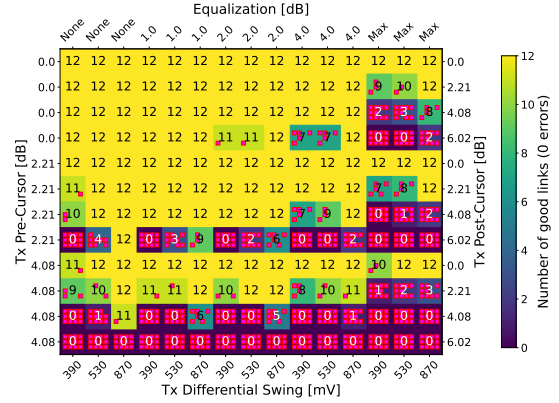
(c) ICjul22c.



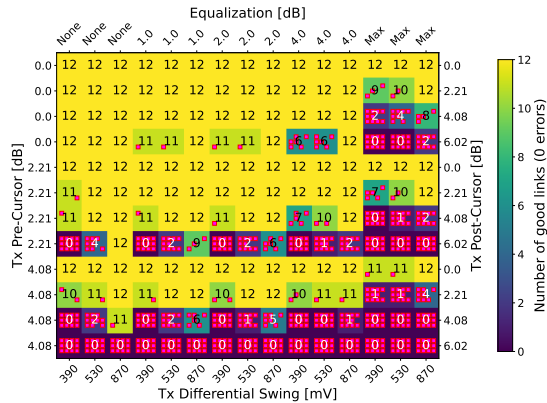
(d) ICjul22d.



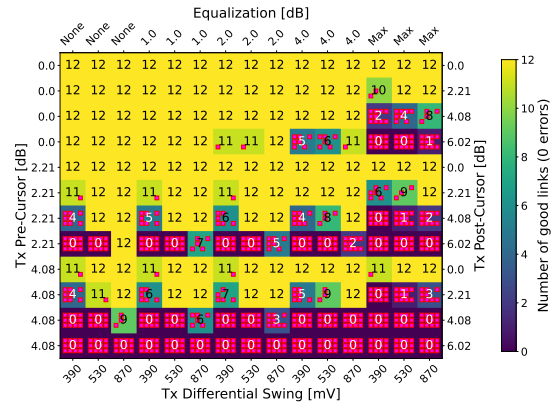
(e) ICjul22e.



(f) ICjul22f.



(g) ICjul22g.



(h) ICjul22h.

Figure B.1: Parameter scan results of the transmitter settings for the 12-channel Beta parts. The scans were performed on Serenity HNZN.



Figure B.2: Bathtub scans for the 12-channel Beta parts with the following well-performing Tx settings: pre-cursor of 4.08 dB, post-cursor of 0 dB, differential swing of 1040 mV, and equalization of 3.7 dB. The stars correspond to the points of measurement, and the dashed lines are the (complementary) error function fits of the data. An estimation of the open area at a BER value of 10^{-12} for each scan is also given. Open areas $< 30\%$ are shown in red. The scans were performed on Serenity HNZN.

B.1.1 Clock and Data Recovery Issue

To verify the CDR issue discussed in Section 5.4.3, part ICjul22d was also tested with the CDR turned off on Serenity HNZN. Figures B.3-B.4 show that the problematic link performs well when CDR is off, and that the CDR off has a negligible effect on the overall performance for the other links. These results are consistent with what was measured on Serenity HNZN.

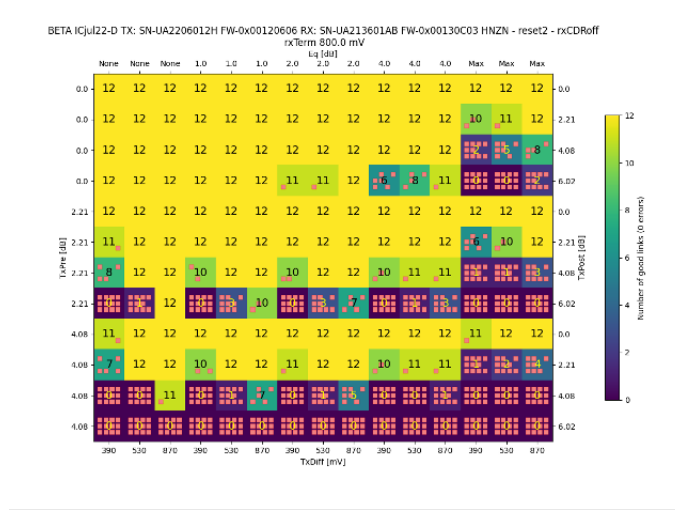


Figure B.3: Parameter scan results for ICjul22d with CDR turned off. The scans were performed on Serenity HNZN.

ZN-beta-10jul22-D-rxCdRoff-sweep17 Bathtub Summary for: h

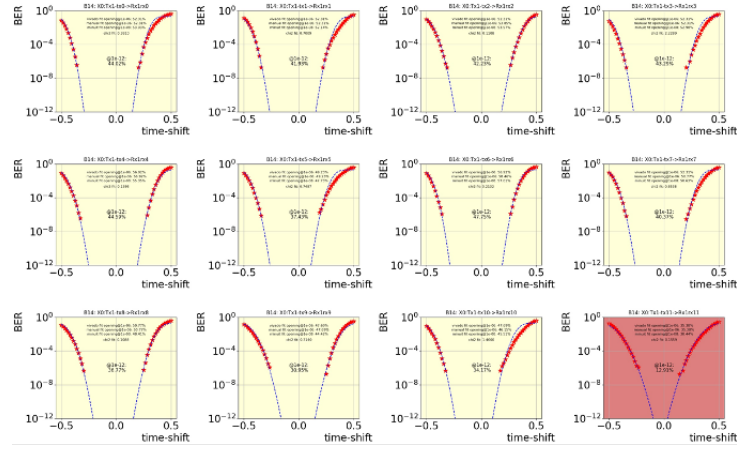


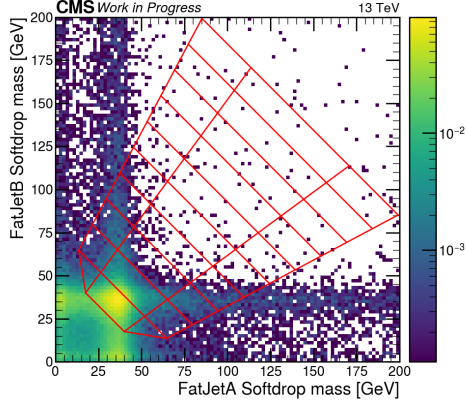
Figure B.4: Bathtub scan for ICjul22d with the CDR turned off for the settings: pre-cursor 4.08 dB, post-cursor 0 dB, differential swing 1040 mV, and equalisation 3.7 dB. The stars correspond to the points of measurement, and the dashed lines are the (complementary) error function fits of the data. An estimation of the open area at a BER value of 10^{-12} is also given. The measurements were performed on Serenity HNZN.

Appendix C

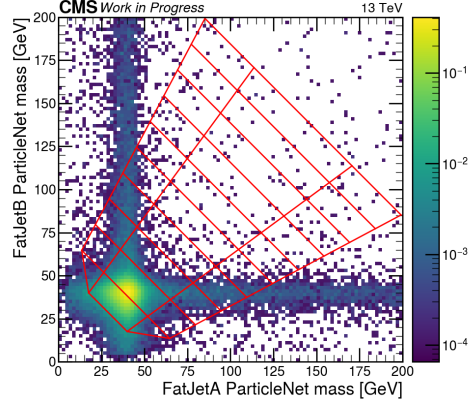
Additional NMSSM Results

C.1 Mass Distribution of Signal Samples

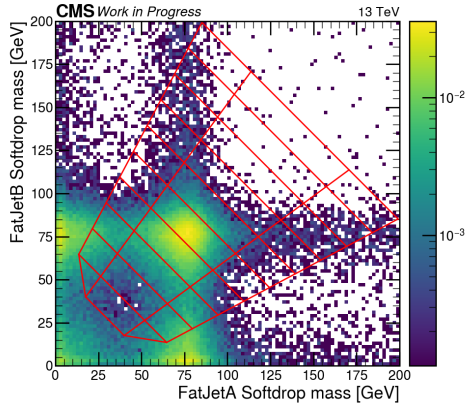
The mass distributions of the selected AK8 jets before any double- b -tag specific cuts are shown in Figure 7.1, for some of the signal samples with varying m_{H_1} . The mass distributions for signal samples can be seen in Figure 7.2. The Softdrop mass grid was used in all cases.



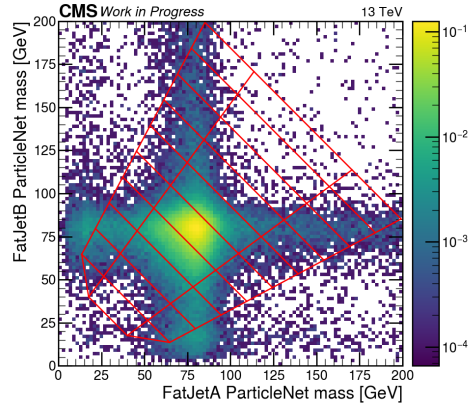
(a) Softdrop, signal sample with $m_{H_1} = 40$ GeV, $m_{\text{SUSY}} = 2800$ GeV.



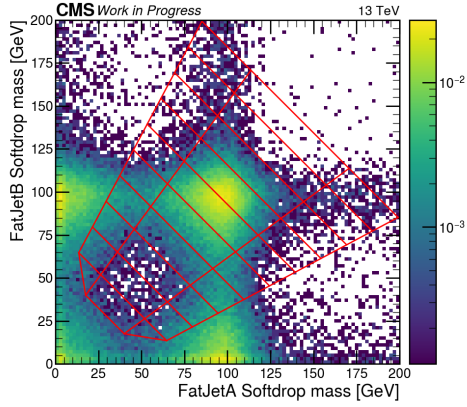
(b) ParticleNet, signal sample with $m_{H_1} = 40$ GeV, $m_{\text{SUSY}} = 2800$ GeV.



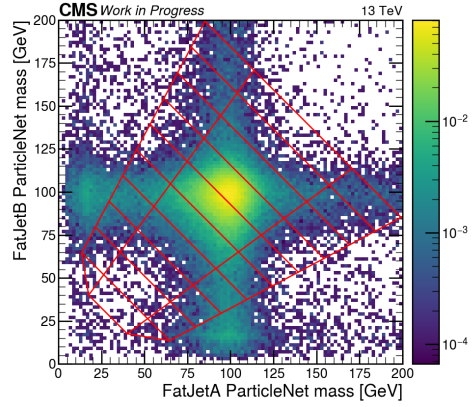
(c) Softdrop, signal sample with $m_{H_1} = 80$ GeV, $m_{\text{SUSY}} = 2800$ GeV.



(d) ParticleNet, signal sample with $m_{H_1} = 80$ GeV, $m_{\text{SUSY}} = 2800$ GeV.

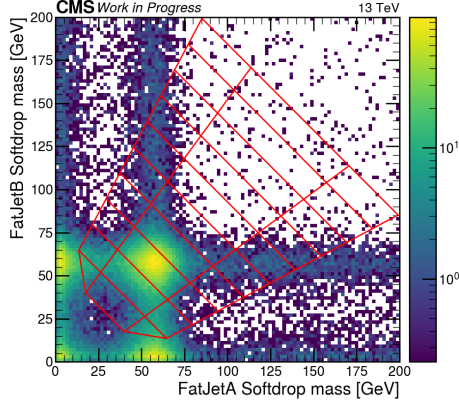


(e) Softdrop, signal sample with $m_{H_1} = 100$ GeV, $m_{\text{SUSY}} = 2800$ GeV.

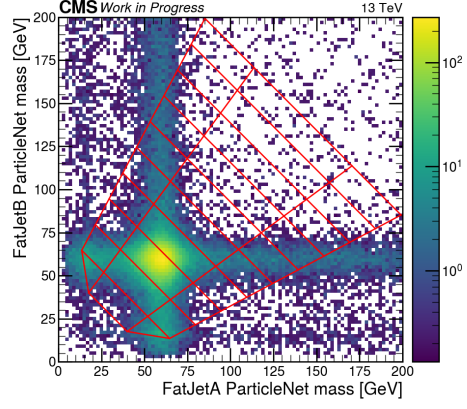


(f) ParticleNet, signal sample with $m_{H_1} = 100$ GeV, $m_{\text{SUSY}} = 2800$ GeV.

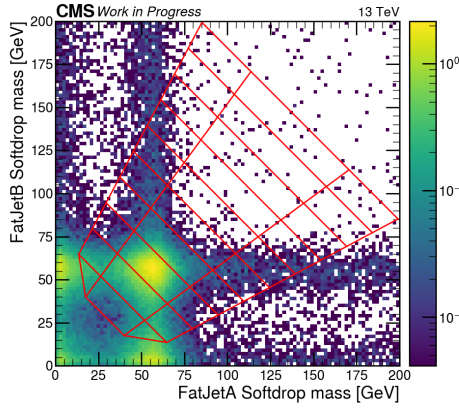
Figure 7.1: Distribution of the reconstructed jet mass for the two selected AK8 jets for various m_{H_1} masses for the Softdrop algorithm (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection were included.



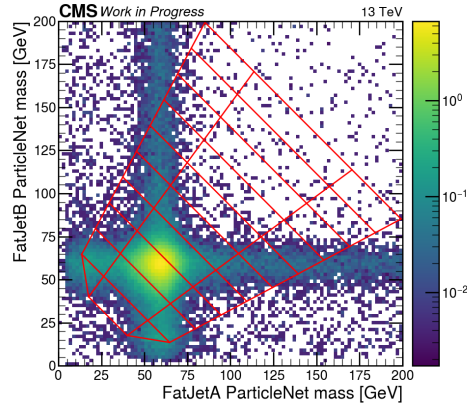
(a) Softdrop, signal sample with $m_{H_1} = 60$ GeV, $m_{\text{SUSY}} = 1200$ GeV.



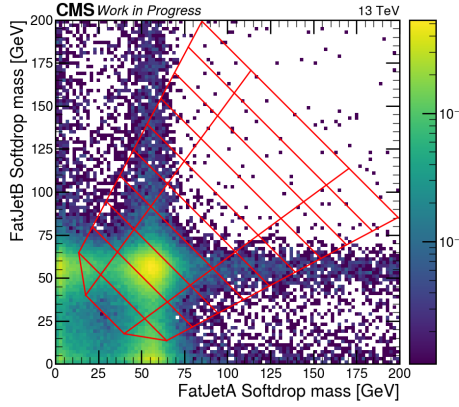
(b) ParticleNet, signal sample with $m_{H_1} = 60$ GeV, $m_{\text{SUSY}} = 1200$ GeV.



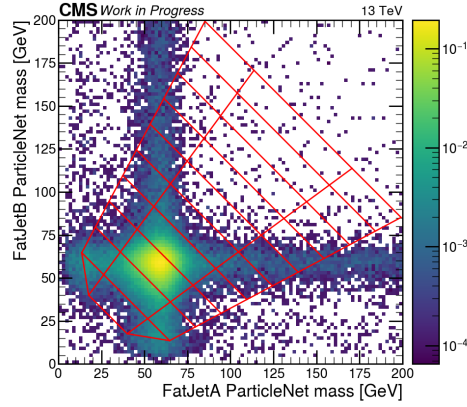
(c) Softdrop, signal sample with $m_{H_1} = 60$ GeV, $m_{\text{SUSY}} = 2000$ GeV.



(d) ParticleNet, signal sample with $m_{H_1} = 60$ GeV, $m_{\text{SUSY}} = 2000$ GeV.



(e) Softdrop, signal sample with $m_{H_1} = 60$ GeV, $m_{\text{SUSY}} = 2800$ GeV.



(f) ParticleNet, signal sample with $m_{H_1} = 60$ GeV, $m_{\text{SUSY}} = 2800$ GeV.

Figure 7.2: Distribution of the reconstructed jet mass for the two selected AK8 jets for various m_{SUSY} masses for the Softdrop algorithm (left) and the ParticleNet algorithm (right). All events that passed the kinematic event selection were included.

Acronyms

2D two-dimensional.

2S Two-Strip.

AK anti- k_T .

ASIC Application-Specific Integrated System.

ATCA Advanced Telecom Computer Architecture.

AUC Area Under the Curve.

BDT Boosted Decision Tree.

BER Bit Error Ratio.

BRAM Block RAM.

BSM Beyond the Standard Model.

CDR Clock and Data Recovery.

CLB Configurable Logic Block.

CMS Compact Muon Solenoid.

CPU Central Processing Unit.

CR Control Region.

CSC Cathode Strip Chamber.

DAQ Data Acquisition.

DSP Digital Signal Processor.

DT Drift Tube chamber.

DTC Data Trigger and Control.

ECAL Electromagnetic Calorimeter.

FIFO First In First Out.

FPGA Field Programmable Gate Array.

GEM Gas Electron Multiplier chamber.

GT Global Trigger.

GUT Grand Unified Theory.

HCAL Hadron Calorimeter.

HDL Hardware Description Language.

HF Hadron Forward.

HGCAL High Granularity Calorimeter.

HL-LHC High-Luminosity LHC.

HLS High-Level Synthesis.

HLT High-Level Trigger.

HO Hadron Outer.

I/O Input/Output.

IBERT Integrated Bit Error Ratio Tester.

IP Intellectual Property.

IR Input Router.

ISI Inter-Symbol Interference.

JER Jet Energy Resolution.

JES Jet Energy Scale.

JMR Jet Mass Resolution.

JMS Jet Mass Scale.

L1 Level-1.

L1A Level-1 Accept.

L1T Level-1 Trigger.

LHC Large Hadron Collider.

lpGBT Low Power GigaBit Transceiver.

LSP Lightest SUSY Particle.

LUT Look-Up Table.

MC Monte Carlo.

MC Match Calculator.

ME Match Engine.

MIP Minimum Ionizing Particle.

MP Match Processor.

MSSM Minimal Supersymmetric Standard Model.

MTD MIP Timing Detector.

NLSP Next-to-LSP.

NMSSM Next-to-Minimal Supersymmetric Standard Model.

OMA Optical Modulation Amplitude.

PF Particle Flow.

PR Projection Router.

PRBS Pseudorandom Binary Sequence.

PS Proton Synchrotron.

PS Pixel-Strip.

PUPPI PileUp Per Particle Identification.

QCD Quantum Chromodynamics.

RAM Random-Access Memory.

RMS Root Mean Square.

ROC Receiver Operating Characteristic.

RPC Resistive Plate Chamber.

RSSI Received Signal Strength Indicator.

Rx Receiver.

SM Standard Model.

SMASH Serenity MAnagement SHell.

SoC System-on-Chip.

SPS Super Proton Synchrotron.

SUSY Supersymmetry.

TB Track Builder.

TC Tracklet Calculator.

TE Tracklet Engine.

TF Track Finder.

TM Track Merger.

TP Tracklet Processor.

TR Tag Region.

Tx Transmitter.

URAM UltraRAM.

VCSEL Vertical-Cavity Surface-Emitting Laser.

VHDL Very High-Speed Integrated Circuit Hardware Description Language.

VM Virtual Module.

VMR VMRouter.

VR Validation Region.

Bibliography

- [1] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (Sept. 2012) 30–61 (cit. on pp. 1, 4).
- [2] *Standard Model of Elementary Particles*, accessed 2022-07-19, URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (cit. on p. 2).
- [3] R. Workman et al., *Review of Particle Physics*, (2022) (cit. on pp. 2, 7).
- [4] B. R. Martin, *Nuclear and Particle Physics: An Introduction*, 2nd ed., John Wiley and Sons Ltd, 2009, ISBN: 9780470742754 (cit. on p. 2).
- [5] M. Thomson, *Modern particle physics*, New York: Cambridge University Press, 2013, ISBN: 978-1-107-03426-6 (cit. on p. 2).
- [6] W. Pauli, *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*, Zeitschrift für Physik **31** (1925) 765–783 (cit. on p. 2).
- [7] S. Khalil and S. Moretti, *Supersymmetry Beyond Minimality: From Theory to Experiment*, CRC Press, 2017, ISBN: 9781498756747 (cit. on p. 6).
- [8] D. Contardo et al., *Technical Proposal for the Phase-II Upgrade of the CMS Detector*, tech. rep., Geneva, June 2015, URL: <https://cds.cern.ch/record/2020886> (cit. on pp. 6, 13, 27, 28, 31).
- [9] CMS Collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, Nature **607** (July 2022) 60–68 (cit. on p. 6).
- [10] U. Ellwanger and A. M. Teixeira, *NMSSM with a singlino LSP: possible challenges for searches for supersymmetry at the LHC*, Journal of High Energy Physics **2014** (Oct. 2014) (cit. on pp. 7, 9, 10, 101, 105, 107).
- [11] U. Ellwanger, C. Hugonie and A. M. Teixeira, *The Next-to-Minimal Supersymmetric Standard Model*, Physics Reports **496** (Nov. 2010) 1–77 (cit. on p. 9).

- [12] CMS Collaboration, *Search for light Higgs bosons from supersymmetric cascade decays in pp collisions at $\sqrt{s} = 13$ TeV*, tech. rep., Geneva: CERN, 2022, arXiv: 2204.13532, URL: <https://cds.cern.ch/record/2807853> (cit. on pp. 10, 102, 105–107, 109, 110, 117, 126, 129, 132).
- [13] G. Apollinari et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs, Geneva: CERN, 2017, URL: <https://cds.cern.ch/record/2284929> (cit. on pp. 11, 26).
- [14] L. R. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001, URL: <https://cds.cern.ch/record/1129806> (cit. on pp. 12, 15).
- [15] CERN, *The third run of the Large Hadron Collider has successfully started*, accessed 2022-07-19, July 2022, URL: <https://home.cern/news/news/cern/third-run-large-hadron-collider-has-successfully-started> (cit. on p. 12).
- [16] Service graphique CERN, *Overall view of the LHC. Vue d'ensemble du LHC*, (2014), URL: <http://cds.cern.ch/record/1708849> (cit. on p. 13).
- [17] CMS Collaboration, *Public CMS Luminosity Information*, accessed 2022-09-11, URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults> (cit. on p. 14).
- [18] CERN, *LHC Report: Protons: mission accomplished*, accessed 2022-07-19, Oct. 2018, URL: <https://home.cern/news/news/physics/lhc-report-protons-mission-accomplished> (cit. on p. 14).
- [19] *Longer term LHC schedule*, accessed 2022-07-19, Jan. 2022, URL: <http://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm> (cit. on p. 14).
- [20] S. Fartoukh et al., *LHC Configuration and Operational Scenario for Run 3*, tech. rep., Geneva: CERN, Nov. 2021, URL: <https://cds.cern.ch/record/2790409> (cit. on p. 14).
- [21] G. L. Bayatian, S. Chatrchyan and G. Hmayakyan, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, Technical design report. CMS, Geneva: CERN, 2006, URL: <https://cds.cern.ch/record/922757> (cit. on pp. 15, 19, 22).
- [22] T. Sakuma, *Cutaway diagrams of CMS detector*, (2019), URL: <https://cds.cern.ch/record/2665537> (cit. on p. 16).
- [23] O. Kodolova and M. Murray, *Heavy Ion Physics with CMS*, tech. rep., Geneva: CERN, 2009, arXiv: 0907.4492, URL: <https://cds.cern.ch/record/1204545> (cit. on p. 17).

- [24] CMS Collaboration, *The Phase-2 Upgrade of the CMS Tracker*, tech. rep., Geneva: CERN, June 2017, URL: <https://cds.cern.ch/record/2272264> (cit. on pp. 18, 27, 29).
- [25] A. Dominguez et al., *CMS Technical Design Report for the Pixel Detector Upgrade*, tech. rep., 2012, URL: <http://cds.cern.ch/record/1481838> (cit. on p. 19).
- [26] CMS Collaboration, *The Phase-2 Upgrade of the CMS DAQ Interim Technical Design Report*, tech. rep., Geneva: CERN, 2017, URL: <https://cds.cern.ch/record/2283193> (cit. on p. 21).
- [27] A. M. Sirunyan et al., *Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JINST **15** (2020) P10017, arXiv: 2006.10165 [hep-ex] (cit. on p. 21).
- [28] V. Khachatryan et al., *The CMS trigger system*, JINST **12** (2017) P01020, arXiv: 1609.02366 [physics.ins-det] (cit. on p. 21).
- [29] CMS Collaboration, *The Phase-2 Upgrade of the CMS L1 Trigger Interim Technical Design Report*, tech. rep. CERN-LHCC-2017-013. CMS-TDR-017, Geneva: CERN, Sept. 2017, URL: <https://cds.cern.ch/record/2283192> (cit. on pp. 22, 29, 33, 37, 38, 60, 61).
- [30] A. M. Sirunyan et al., *Particle-flow reconstruction and global event description with the CMS detector*, JINST **12** (2017) P10003, arXiv: 1706.04965 [physics.ins-det] (cit. on p. 23).
- [31] D. Bertolini et al., *Pileup Per Particle Identification*, JHEP **10** (2014) 059, arXiv: 1407.6013 [hep-ph] (cit. on pp. 23, 24).
- [32] A. M. Sirunyan et al., *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}=13$ TeV*, Journal of Instrumentation **13** (June 2018) P06015 (cit. on p. 23).
- [33] A. M. Sirunyan et al., *Pileup mitigation at CMS in 13 TeV data*, JINST **15** (2020) P09018, arXiv: 2003.00503 [hep-ex] (cit. on p. 24).
- [34] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ex] (cit. on p. 24).
- [35] A. M. Sirunyan et al., *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, Journal of Instrumentation **13** (May 2018) P05011 (cit. on pp. 24, 25, 102).
- [36] *The HL-LHC project*, accessed 2023-01-01, URL: <https://hilumilhc.web.cern.ch/content/hl-lhc-project> (cit. on pp. 26, 27).
- [37] CMS Collaboration, *A MIP Timing Detector for the CMS Phase-2 Upgrade*, tech. rep., Geneva: CERN, 2019, URL: <https://cds.cern.ch/record/2667167> (cit. on p. 28).

- [38] CMS Collaboration, *The Phase-2 Upgrade of the CMS Barrel Calorimeters*, tech. rep., Geneva: CERN, 2017, URL: <https://cds.cern.ch/record/2283187> (cit. on p. 28).
- [39] CMS Collaboration, *The Phase-2 Upgrade of the CMS Endcap Calorimeter*, tech. rep., Geneva: CERN, 2017, URL: <https://cds.cern.ch/record/2293646> (cit. on p. 29).
- [40] CMS Collaboration, *The Phase-2 Upgrade of the CMS Muon Detectors*, tech. rep., Geneva: CERN, 2017, URL: <https://cds.cern.ch/record/2283189> (cit. on p. 29).
- [41] CMS Collaboration, *The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger*, tech. rep., Geneva: CERN, 2021, URL: <https://cds.cern.ch/record/2759072> (cit. on p. 29).
- [42] R. Aggleton et al., *An FPGA based track finder for the L1 trigger of the CMS experiment at the High Luminosity LHC*, Journal of Instrumentation **12** (Dec. 2017) P12019–P12019 (cit. on pp. 32, 35, 36, 60).
- [43] P. Moreira et al., *lpGBT documentation: release*, 2022, URL: <https://cds.cern.ch/record/2809058> (cit. on p. 32).
- [44] *AdvancedTCA Base Specification*, ECN 3.0-3.0-001, Rev. 3.0, PICMG, Apr. 2015, URL: https://www.picmg.org/wp-content/uploads/ECN001_PICMG_ATCA_3_0_R3_0-RELEASED-2015-04-20b.pdf (cit. on p. 33).
- [45] *Serenity*, accessed 12-2022, 2022, URL: <https://serenity.web.cern.ch/serenity> (cit. on p. 33).
- [46] *UltraScale™ Architecture and Product Data Sheet: Overview*, DS890, Rev. 4.3, Xilinx, Nov. 2022, URL: <https://docs.xilinx.com/v/u/en-US/ds890-ultrascale-overview> (cit. on pp. 33, 39, 40).
- [47] *The Apollo Platform*, accessed 12-2022, 2022, URL: apollo-blade.info (cit. on p. 34).
- [48] A. Albert et al., *The Apollo ATCA design for the CMS track finder and the pixel readout at the HL-LHC*, JINST **17** (Apr. 2022) C04033 (cit. on p. 34).
- [49] E. Bartz et al., *FPGA-based tracking for the CMS Level-1 trigger using the tracklet algorithm*, 2019, arXiv: 1910.09970 (cit. on pp. 35–37, 42).
- [50] A. Hart, *Level 1 Track Finder at CMS*, tech. rep., 2019, arXiv: 1910.06614, URL: <https://cds.cern.ch/record/2693909> (cit. on p. 38).
- [51] S. D. Brown et al., *Field-Programmable Gate Array Technology*, Springer, 1992, ISBN: 978-1-4615-3572-0 (cit. on p. 39).
- [52] IEEE Computer Society, *IEEE Standard VHDL Language Reference Manual*, IEEE Std 1076-2008, IEEE, 2009 (cit. on p. 40).

- [53] *Vivado Design Suite User Guide: High-Level Synthesis*, UG902, v2020.1, Xilinx, May 2021, URL: <https://docs.xilinx.com/v/u/en-US/ug902-vivado-high-level-synthesis> (cit. on p. 40).
- [54] B. Stroustrup, *The C++ Programming Language*, 4th, Addison-Wesley, 2013, ISBN: 0321563840 (cit. on p. 40).
- [55] M.-L. Holmberg, *The Level-1 Track Finder for the CMS High-Luminosity LHC Upgrade*, tech. rep., Geneva: CERN, 2022, URL: <https://cds.cern.ch/record/2844896> (cit. on p. 55).
- [56] CMS Collaboration, *The Phase-2 Upgrade of the CMS L1 Trigger Interim Technical Design Report*, tech. rep., Geneva: CERN, 2017, URL: <https://cds.cern.ch/record/2283192> (cit. on p. 60).
- [57] Stanislava Sevova, *PF and PUPPI Algo Update*, June 2017, URL: https://indico.cern.ch/event/664087/contributions/2711604/attachments/1519483/2373066/FastPUPPI_P2L1TrigUpgMtg_090617_ssevova.pdf (cit. on p. 61).
- [58] *FireFly™ Active Optical Micro Flyover System™ Cable Assembly*, accessed 2023-04-05, Samtec, URL: <https://www.samtec.com/products/ecuo> (cit. on pp. 68, 69).
- [59] H. J. R. Dutton, *Understanding Optical Communications*, IBM, 1998 (cit. on pp. 69, 92).
- [60] *MTP® Connectors*, accessed 2023-04-28, US Conec Ltd., URL: <https://www.usconec.com/connectors/mtp-connectors> (cit. on p. 69).
- [61] C. W. Wilmsen, H. Temkin and L. A. Coldren, *Vertical-Cavity Surface-Emitting Lasers: Design, Fabrication, Characterization, and Applications*, Cambridge University Press, 2001, ISBN: 9780521006293 (cit. on p. 69).
- [62] *SMASH: The Serenity Management Shell*, accessed 2023-04-05, URL: <https://serenity.web.cern.ch/serenity/smash/> (cit. on p. 73).
- [63] *An Attribute-Programmable PRBS Generator and Checker*, accessed 2023-04-05, URL: https://docs.xilinx.com/v/u/en-US/xapp884_PRBS_GeneratorChecker (cit. on p. 74).
- [64] A. Athavale and C. Christensen, *High-Speed Serial I/O Made Simple*, Xilinx, 2005, URL: <https://www.xilinx.com/publications/archives/books/serialio.pdf> (cit. on p. 74).
- [65] *Optical Modulation Amplitude (OMA) and Extinction Ratio*, Maxim Integrated, 2009, URL: <https://pdfserv.maximintegrated.com/en/an/AN2710.pdf> (cit. on p. 96).
- [66] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Physics Reports **516** (July 2012) 1–102 (cit. on p. 101).

- [67] A. J. Larkoski et al., *Soft drop*, Journal of High Energy Physics **2014** (May 2014) (cit. on pp. 102, 104, 116).
- [68] CMS Collaboration, *Identification of double-b quark jets in boosted event topologies*, tech. rep., Geneva: CERN, 2016, URL: <https://cds.cern.ch/record/2195743> (cit. on pp. 102, 103).
- [69] H. Qu and L. Gouskos, *Jet tagging via particle clouds*, Physical Review D **101** (Mar. 2020) (cit. on pp. 102, 103).
- [70] Y. Dokshitzer et al., *Better jet clustering algorithms*, Journal of High Energy Physics **1997** (Aug. 1997) 001–001 (cit. on p. 104).
- [71] CMS Collaboration, *Mass regression of highly-boosted jets using graph neural networks*, (2021), URL: <https://cds.cern.ch/record/2777006> (cit. on pp. 105, 116).
- [72] W. Beenakker et al., *NNLL-fast: predictions for coloured supersymmetric particle production at the LHC with threshold and Coulomb resummation*, Journal of High Energy Physics **2016** (Dec. 2016) (cit. on p. 106).
- [73] A. Djouadi, J. Kalinowski and M. Spira, *HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Computer Physics Communications **108** (Jan. 1998) 56–74 (cit. on p. 107).
- [74] A. Djouadi et al., *HDECAY: Twenty++ years after*, Computer Physics Communications **238** (May 2019) 214–231 (cit. on p. 107).
- [75] F. Déliot et al., *Properties of the Top Quark*, Annual Review of Nuclear and Particle Science **64** (2014) 363–381 (cit. on p. 108).
- [76] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, Journal of High Energy Physics **2014** (July 2014) (cit. on p. 108).
- [77] R. Frederix et al., *The automation of next-to-leading order electroweak calculations*, Journal of High Energy Physics **2018** (July 2018) (cit. on p. 108).
- [78] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: [hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146) (cit. on p. 108).
- [79] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: [0709.2092](https://arxiv.org/abs/0709.2092) [hep-ph] (cit. on p. 108).
- [80] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: [1002.2581](https://arxiv.org/abs/1002.2581) [hep-ph] (cit. on p. 108).
- [81] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159–177, ISSN: 0010-4655 (cit. on p. 108).

- [82] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250–303, ISSN: 0168-9002 (cit. on p. 108).
- [83] M. Peruzzi et al., *The NanoAOD event data format in CMS*, Journal of Physics: Conference Series **1525** (Apr. 2020) 012038 (cit. on p. 108).
- [84] *The NanoAOD-Tools*, accessed 2023-05-10, URL: <https://twiki.cern.ch/twiki/bin/view/Sandbox/NanoAODToolsTMP> (cit. on p. 108).
- [85] *nanoAOD-tools*, accessed 2023-05-10, URL: <https://github.com/cms-nanoAOD/nanoAOD-tools> (cit. on p. 108).
- [86] T. Adams et al., *Beam test evaluation of electromagnetic calorimeter modules made from proton-damaged PbWO₄ crystals*, Journal of Instrumentation **11** (Apr. 2016) P04012 (cit. on p. 109).
- [87] J. Taylor, *A search for light Higgs bosons in supersymmetric decay cascades with the CMS detector*, (Oct. 2019) (cit. on pp. 111, 113).
- [88] W. Buttinger, *Background Estimation with the ABCD Method*, accessed 2023-05-04, Oct. 2018, URL: https://twiki.cern.ch/twiki/pub/Main/ABCDMethod/ABCDGuide_draft18Oct18.pdf (cit. on p. 127).
- [89] CMS Collaboration, *Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks*, (2022), arXiv: 2205.06667 [hep-ex] (cit. on p. 131).
- [90] A. Apresyan et al., *Search for Gluon Fusion Nonresonant HH to 4b using boosted Higgs-tagged AK8 jets*, (Oct. 2020), Internal CMS analysis note., eprint: AN-2020/201v17, URL: https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2020/201 (cit. on p. 131).
- [91] D. Brehm et al., *Search for nonresonant Higgs boson pair production via vector boson fusion in boosted 4b final state at $\sqrt{s} = 13$ TeV*, (Dec. 2020), Internal CMS analysis note., eprint: AN-2020/231v8, URL: https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2020/231 (cit. on p. 132).
- [92] P. Bortignon et al., *Calibration of boosted jet flavour taggers in full Run 2 data*, (Jan. 2021), Internal CMS analysis note., eprint: AN-2021/005v10, URL: https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2021/005 (cit. on p. 132).
- [93] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** (Feb. 2011) (cit. on p. 135).
- [94] *Combine*, accessed 2023-05-10, URL: <http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/> (cit. on p. 136).