Search for Higgs boson pair production and study of the high granularity timing detector peripheral electronics in the ATLAS experiment

by

Han Liangliang[1]

Directed by

Professor Chen Shenjian[1]

[1]School of Physics, Nanjing University

May,2024

Submitted in partial fulfillment of the requirements
for the degree of PhD in Particle Physics and Nuclear Physics

南京大学研究生毕业论文英文摘要首页用纸

THESIS: Search for Higgs boson pair production and study of the high granularity timing detector peripheral electronics in the ATLAS experiment
SPECIALIZATION: Particle Physics and Nuclear Physics
POSTGRADUATE: Han Liangliang
MENTOR: Professor Chen Shenjian

Within the framework of the Standard Model, the Higgs mechanism explains how fundamental particles acquire mass. It achieves this by introducing a self-coupling mechanism for the Higgs boson, allowing the Higgs field to obtain a non-zero vacuum expectation value, thereby leading to the breaking of electroweak symmetry and enabling other particles to acquire mass. Measuring the self-coupling of the Higgs boson provides us with a better understanding of the shape of the Higgs potential, facilitating further investigation into significant issues such as testing the Higgs mechanism, early universe thermodynamic processes, and the stability of the cosmic vacuum. At the Large Hadron Collider (LHC), studying Higgs boson pair production (HH) allows for the measurement of the Higgs boson self-coupling. Additionally, the presence of anomalous couplings, as proposed by many new physics hypotheses beyond the Standard Model, can alter the Higgs potential and its self-coupling strength, thereby increasing the production rate of HH events. It is possible to discover Higgs boson pair production within the data collected thus far. Exploring Higgs boson pair production is a significant research focus of the ATLAS experiment, which has the potential to deepen our fundamental understanding of the Higgs boson potential and its self-coupling.

The first part of this thesis is to search for non-resonant production of Higgs boson pairs in the final state of two b–jets and two τ-leptons ($b\bar{b}\tau^+\tau^-$) through the HH ggF and VBF production modes by using the 140 fb^{-1} Run-2 data collected by the ATLAS detector at the centreofmass energy of 13 TeV, thus the Higgs boson self-coupling modifier κ_λ can be measured. This analysis gives the best measurement of this quantity in the HH final state. This search is performed based on the previous round of the analysis. Notable improvements in experiment sensitivity are achieved through various updates, including more precise study of the key backgrounds, the inclusion of dedicated VBF SR, improvement of BDT strategy, and etc. No significant excess above the expected background from Standard Model processes is observed. The observed (expected) 95% confidence level (CL) upper limit on μ_{HH} is 5.9 (3.3) times the Standard Model prediction. The observed limit on μ_{HH} is looser than the expected one as a result of a mild excess in the high-mHH region of $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel. The corresponding observed (expected) 95% confidence intervals for the self-coupling modi-

fier κ_λ and the quartic coupling modifier κ_{2V} are respectively $-3.1 < \kappa_\lambda < 9.0$ ($-2.5 < \kappa_\lambda < 9.3$) and $-0.5 < \kappa_{2V} < 2.7$ ($-0.2 < \kappa_{2V} < 2.4$).

Although we have achieved quite good measurement of the Higgs boson self-coupling, there is still room for improvement towards more precise regime. This goal can be realized by collecting more collision data. The HL-LHC project is established, targeting for the extension of LHC operability and an increase of the collision rate, which means the ability of producing much more HH events. In order to deal with the challenges posed by the HL-LHC, especially the adverse effects of the pile-up, the ATLAS detector has to be upgraded (called ATLAS Phase-2 upgrade) to properly handle the flood of incoming data and increased irradiation level. A high granularity timing detector (HGTD) is therefore proposed for the ATLAS upgrade. With the timing information provided by the HGTD, the tracks that are spacially overlapped can still be distinguished. The PEB board plays an important role in the HGTD peripheral electronics system, and it serves as a bridge that connects the HGTD modules and the off-detector systems. Its design faces many difficulties, such as high complexity, huge data throughput, working in an irradiation environment, and limited installation space.

The second part of this thesis is to design a prototype before the final PEB design. The design requirement is: under the circumstances of key chips scarcity, all key functions of PEB should be integrated into the prototype for verification to ensure the normal progress of HGTD peripheral electronics research and development. It aims to verify the following five aspects: system configuration, data transmission, power supply distribution, monitoring network, clock and fast command distribution. Among them, the verification of data transmission is more complicated and needs to consider versatile link (lpGBT + VTRx+) communication, full data path communication, E-port data rate setting, etc. In response to the design requirement raised above, it was decided to adopt a modular design, that is, to integrate the key chips onto pluggable daughter boards. This design allows key chips to be replaced, and even if a key chip fails, it will not cause the waste of other chips in the system. Functionally, this design is a simplified PEB, but it can achieve all the above functions to be verified by using all kinds of key chips. Based on the joint test system, we tested the above-mentioned key functions one by one, which all meet the specified requirements and work as expected. The modular PEB demonstrates full operational capability in terms of hardware design and verifies the feasibility of key functions, which gives us more confidence on the future PEB board design. Certainly, the modular PEB can not only be used as a PEB-related electronics test and verification platform, but has also been used by many HGTD sub-groups (such as CERN, Nikhef and KTH) for different tasks, greatly accelerating the progress of these tasks.

KEYWORDS: ATLAS experiment; Higgs boson pair production; Higgs boson self coupling;

High granularity timing detector; Peripheral electronics board

Contents

List of Figures

List of Tables

Introduction

Since the discovery of the Higgs boson [1, 2] at the Large Hadron Collider in 2012, a priority of the ATLAS and CMS collaborations has been to better understand its properties and couplings. An important test of the electroweak symmetry breaking would be to establish evidence of the Higgs boson self-coupling (λ_{HHH}), which is directly linked to the shape of the Higgs potential. This can be achieved through a measurement of Higgs boson pair (HH) production. The LHC is the only large-scale scientific facility in the world that can perform such kind of measurement. The two LHC experiments, ATLAS and CMS, has been dedicated to the search for HH production since the first data taking (Run-1).

In the Standard Model, Higgs boson pair events are dominantly produced via the gluon-gluon fusion (ggF) processes at the LHC, involving both the Yukawa coupling to top quarks and the Higgs boson self-coupling, so-called box and triangle processes. These two processes interfere destructively, leading to a very small HH cross-section. The second most dominant mode to produce a Higgs boson pair in the SM is vector-boson fusion (VBF) with an even smaller cross-section. Deviations of the Higgs boson couplings from the SM predictions could lead to a significant enhancement of the Higgs boson pair production rate. The Higgs boson trilinear self-coupling modifier due to BSM scenarios [3] is defined as $\kappa_\lambda = \lambda_{HHH}/\lambda^{SM}_{HHH}$. For example, at $\kappa_\lambda = 0$, the HH production rate doubles, and at $\kappa_\lambda = -1$ the rate quadruples [4]. The modifier κ_{2V} of the quartic $HHVV(V = W, Z)$ coupling is defined in the same way. The Variations of κ_{2V} would also greatly enhance the VBF HH production, leading to an increase of two orders of magnitude at coupling strength twice the SM prediction [5].

Figure 1 shows the HH decay modes that are considerable at both the ATLAS and CMS experiments. Those decay modes are separately searched at the beginning, then combined together to achieve the best sensitivity with the collected data. Based on the branching ratio and final state cleaness of the decay modes, the three most sensitive decay modes (analysis channels) are respectively $b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$ and $b\bar{b}\gamma\gamma$.

Both ATLAS and CMS experiments have already made a big effort on the measurement of the Higgs boson self-coupling through the process of Higgs boson pair production. For example, the ATLAS experiment has established dedicated analysis teams for those channels with rather promising sensitivities using the increasing data collected along the LHC operation, such as the Run-1 and Run-2 data that have been analyzed, the Run-3 data that

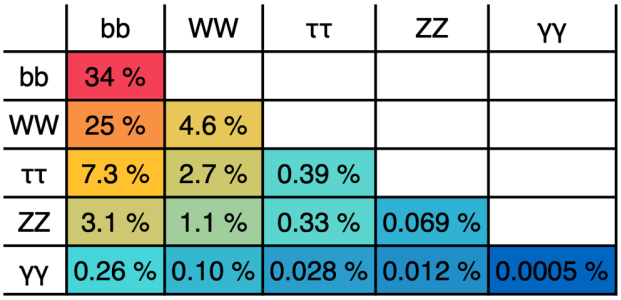


Figure 1: Branching ratios for the most important HH decay modes assuming SM couplings, calculated at NLO [6].

Table 1: Observed and expected 95% CL upper limit on HH signal strength from the $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$ and their statistical combination based on the Run-2 data [7].

	Obs.	-2σ	-1σ	Exp.	1σ	2σ
$b\bar{b}\gamma\gamma$	4.3	3.1	4.1	5.7	8.8	14.3
$b\bar{b}\tau^+\tau^-$	4.6	2.1	2.8	3.9	5.9	9.4
Combined	3.1	1.7	2.2	3.1	4.7	7.3

has been taken since the year of 2023 and analyzed at the same time, the Run-4 and Run-5 data that will be collected during the HL-LHC period. The current measurement of HH comes from the analysis of the Run-2 data, where the 95% Confidence Level (CL) upper limit on HH signal strength and constraint on Higgs self-coupling modifier κ_λ are given. Table 1 shows the observed and expected 95% CL upper limit on HH signal strength from the $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$ and their statistical combination based on the Run-2 data. Table 2 shows the observed and expected 95% CL constraint on κ_λ from the $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$ and their statistical combination based on the Run-2 data. For the HH measurement of CMS experiment, since its results are basically comparable with that of ATLAS experiment, they are not described here. Recently, the sensitivity of the $b\bar{b}\tau^+\tau^-$ channel is further improved based on the ATLAS Run-2 data, and the first part of this thesis will present this study where both the 95% CL upper limit on HH signal strength and 95% CL constraint on κ_λ are measured.

Due to the limited amount of dataset recorded during LHC Phase-1, only a rather loose constraint on the Higgs boson self-coupling can be set, which is pretty far away from having a direct measurement of it. The lack of statistics is the main limiting factor that stops us from performing a more decent measurement on the Higgs boson self-coupling. In fact, the dataset expected at the end of LHC Phase-1 is just a small fraction ($<$ 10%) of the dataset

Table 2: Observed and expected 95% CL constraint on κ_λ from the $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$ and their statistical combination based on the Run-2 data [7].

	Obs.	Exp.
$b\bar{b}\gamma\gamma$	$[-1.6, 6.7]$	$[-2.4, 7.7]$
$b\bar{b}\tau^+\tau^-$	$[-2.4, 9.2]$	$[-2.0, 9.0]$
Combined	$[-1.0, 6.6]$	$[-1.2, 7.2]$

planned for the whole LHC lifetime, and another more than 90% of dataset will be collected during the HL-LHC (LHC Phase-2). The HL-LHC will produce approximately 170 million Higgs bosons and 120,000 Higgs-boson pairs on each of the ATLAS and CMS experiments over a period of about 10 years. With the tenfold Higgs boson events of HL-LHC, the Higgs boson self-coupling will be significantly constrained.

The HL-LHC is expected to deliver beam collisions at the start of 2029 with the peak levelled instantaneous luminosity of $5 \times 10^{34} cm^{-2} s^{-1}$, five times the LHC nominal luminosity. The more intense and concentrated beam makes the pile-up (the number of collisions that happen in one bunch crossing) reach to a higher level. Due to the higher requirement on the beam of the accelerator, some equipment of the LHC have to be upgraded. Also the detectors need to be upgraded to cope with the significantly-increased pile-up. For the ATLAS detector, those sub-detectors that are far away from the collision point and will not suffering from the irradiation damage during the HL-LHC operation, such as the LAr calorimeter, tile calorimeter and the muon spectrometer, will remain unchanged, but their electronics have to be upgraded. Other upgrades include installations of the new inner tracking detector (ITk), high granularity timing detector (HGTD) and new muon chambers.

According to the HL-LHC operation plan, the installation of the updated detectors for ATLAS will start from the year of 2026 until 2029. Currently the HGTD detector is under the stage of research and development. It is proposed for mitigating the adverse effects of pile-up by providing the timing information of tracks. One of the most important parts is the peripheral electronics of HGTD, serving as a bridge that connects the front-end modules and the off-detector systems like the data acquisition (DAQ) system, the luminosity system and the detector control system (DCS). The peripheral electronics is implemented at the peripheral area, which will be covered with peripheral electronics boards (PEB) mounted with a large amount of components. The design of the PEBs is extremely challenging since it needs to handle high data throughput, harsh irradiation environment, limited installation space, complicated connectivity of large amount of components, and interfacing with other sub-systems of the HGTD. Therefore, prior to the PEB prototype, a PEB prototype (modular PEB), as a verification platform, has been designed to help us verify some concerns in advance. The development of such a system is a crucial part of the whole PEB design chain. In addition, it can also greatly benifit other tasks of the HGTD, even become an important step on their own schedules. The design and test of the modular PEB system will be presented in Chapter 5 and Chapter 6.

Clarification of personal contributions

The ATLAS experiment is operated by a close collaboration of about 3000 physicists, engineers and students. Usually, the completion of one work involves the efforts from a group

of people. Therefore, it is important to make it clear how I was involved in the two works (the search for Higgs Boson Pair Production + the design and test of the modular PEB) and also other contributions.

Search for Higgs Boson Pair Production in the $b\bar{b}\tau^+\tau^-$ final state

I was the internal note editor of this analysis and made contributions to both the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels. I was in charge of the production of all the new CxAOD samples (such as the V+jets samples and $t\bar{t}$ dilepton samples) and the maintenance of other CxAOD samples for $\tau_{\text{had}}\tau_{\text{had}}$ channel. I worked on the validation of the background estimation strategy in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, including the estimation of the multi-jet with fake-τ_{had} and $t\bar{t}$ with fake-τ_{had}. I worked on the study of the systematic uncertainties in $\tau_{\text{lep}}\tau_{\text{had}}$ channel, especially the derivation and implementation of theoretical uncertainties on the single-top Wt process. I also worked on the statistical analysis, especially for the $\tau_{\text{lep}}\tau_{\text{had}}$ channel.

Design and test of a verification system for HGTD PEB (modular PEB)

I was the main participant of the modular PEB activities and one of the authors of the PEB specification. I contributed to the hardware design of the modular PEB system, including the design of the module emulator, UPL toolkit and TPS56428 daughter board, the design discussion and check of the carrier board and all the other daughter boards. I developed all the firmwares required by the system based on some common firmware cores from the CERN ASIC design groups, including the firmware for module emulator and its DAQ counterpart, and also the firmware for lpGBT emulator and its DAQ counterpart. I performed most of the individual tests and established the joint test system, based on which those crucial items are then tested.

Other contributions to the ATLAS experiment

I performed unit test for the classes of the ATLAS simulation framework as my qualification task in the ATLAS simulation team. I worked on the Top reweighting in the search for singly-produced Leptoquarks decaying to $b\tau$. I took shifts to monitor the operation of the ATLAS Distributed Computing System.

Chapter 1

Theory

This chapter gives a comprehensive introduction to the theoretical aspects of particle physics, emphasizing physics about the Higgs boson within the framework of the LHC. Additionally, the motivation to search for Higgs boson pair production is explained.

1.1 The Standard Model

1.1.1 Quantum chromodynamics and electroweak theory

The theoretical framework known as the Standard Model (SM) [8–10] of particle physics offers a comprehensive explanation of the electromagnetic, weak, and strong interactions, as well as the elementary particles, in a unified manner. Over the past few decades, numerous experiments have extensively tested it. The results overwhelmingly align with the predictions of the SM, making it the most remarkable achievement in the field of particle physics. The Standard Model (SM) is built upon the principles of Quantum Field Theory (QFT), which provides a framework for understanding the behavior of elementary particles and their interactions. It incorporates the local gauge principle, which governs the nature of these interactions. Moreover, the SM encompasses the principles of Electroweak (EW) symmetry breaking that plays a crucial role in imparting mass to particles.

The fundamental interactions are built within a symmetry group of $SU(3)_C \times SU(2)_L \times U(1)_Y$, resulting in a total of twelve vector fields, whose excitations, referred to as gauge bosons, are responsible for transmitting the forces when particles are interacting with each other. The strong interaction is propagated by eight bi-colored gluons (g), while the weak interaction is propagated by the $W^{\pm}$ and Z bosons. The electromagnetic interaction, on the other hand, is propagated by the photon (γ).

The Dirac equation governs the dynamics of spinor fields. The fundamental constituents of the material universe, known as matter particles or fermions, arise from the excitations of these spinor fields. These fermions can be classified into two categories: leptons labelled as $\binom{\nu_e}{e}, \binom{\nu_\mu}{\mu}, \binom{\nu_\tau}{\tau}$ and quarks labelled as $\binom{u}{d}, \binom{c}{s}, \binom{t}{b}$. Leptons do not engage in strong interaction,

while quarks participate in both electromagnetic and strong interactions.

As demonstrated previously, the Standard Model consists of six leptons and six quarks, which are categorized into three generations based on their masses and decays. In order to provide mass to the bosons and fermions, scalar fields are essential in the electroweak symmetry breaking mechanism. The emergence of the Higgs boson (H) is a direct result of this mechanism.

There are two main divisions of interactions, namely the strong interaction and the joint EW interaction. A concise overview of these interactions is provided below. In accordance with the principles of Quantum Field Theory (QFT), the Lagrangian densities [1] are utilized to describe the dynamic and interaction terms of the quantum fields.

1.1.1.1 Strong interaction

The Quantum Chromo-dynamics (QCD) is a gauge theory that describes the strong interaction between quarks and gluons. This theory is invariant under $SU(3)$ local gauge transformation in the colour space. The symmetry group associated with the QCD is known as $SU(3)_C$. The Lagrangian of QCD can be represented in a schematic form as

$$\mathscr{L}_{\mathrm{QCD}} = \bar{Q} i\gamma^{\mu} D_{\mu} Q - \frac{1}{4} G^{a}_{\mu\nu} G^{\mu\nu}_{a}, \tag{1.1}$$

where $Q(x)$ is the quark field. The symbol $Q(x)$ denotes that the quark field Q is dependent on the four-dimensional space-time x. This notation is employed consistently in this chapter. To fulfill the condition of gauge invariance, eight massless vector fields known as the gluon fields G^{a}_{μ} are introduced. They collectively constitute a covariant derivation and a gluon tensor field referred to as

$$\begin{aligned} D_{\mu} &= \partial_{\mu} + i g_S T_a G^{a}_{\mu} \\ G^{a}_{\mu\nu} &= \partial_{\mu} G^{a}_{\nu} - \partial_{\nu} G^{a}_{\mu} - g_S f_{abc} G^{b}_{\mu} G^{c}_{\nu}, \end{aligned} \tag{1.2}$$

The structure constants and generators of $SU(3)_C$ are represented by f_{abc} and $T_a (a = 1 \ldots 8)$ respectively. The Gell-Mann matrices $\frac{\lambda_a}{2}$ are commonly used to represent the generators T_a. The strong interaction constant g_S, also known as the strong interaction strength $\alpha_S = \frac{g_S^2}{4\pi}$, varies with the interaction energy scale. At high energies, α_S becomes sufficiently small that perturbation calculations lose validity. This phenomenon is referred to as asymptotic freedom.

The theory of color confinement prevents the existence of independently occurring quarks and can be attributed to the strong force interactions within the field of Quantum Chromo-dynamics. Quarks are required to exist in color singlet states, also known as colorless states,

[1] The Lagrangian density in the subsequent context is referred to as the Lagrangian for the sake of clarity

which are called hadrons.With high energies, quarks and gluons have the capability to create sprays of hadrons known as jets through the parton shower and subsequent hadronization processes. Aiming for emulating those processes, phenomenological models are applied, extending from the scale of the hard process to hadronization. Parton shower event generators usually adopts these models.

1.1.1.2 Electroweak interaction

The sophisticated and refined joint gauge theory that combines electromagnetic and weak interactions is intricate and beautiful and it is connected to the $SU(2)_L \times U(1)_Y$ symmetry group. The first element of this theory emerges from the $SU(2)$ symmetry inherent in the interaction governed by the weak charged-current. Here, the lower index L denotes that the interaction of the weak charged current solely links with left-handed chiral (LH) particle states and right-handed chiral (RH) antiparticle states. The theory's second aspect integrates the interaction of weak neutral-current with electromagnetic interaction, denoted by subscript Y signifying weak hypercharge. The schematic Lagrangian of the electroweak interaction can be expressed as

$$\mathscr{L}_{\mathrm{EW}} = \bar{L} i\gamma^{\mu} D_{\mu} L + \bar{R} i\gamma^{\mu} D_{\mu} R - \frac{1}{4} W^{a}_{\mu\nu} W_{a}^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{1.3}$$

where the LH spinor (fermion) field, denoted by $L(x) = \frac{1}{2}(1-\gamma^5)F(x)$ with $F(x)$ as the generic fermion field, can be referred to as the following weak isospin doublet

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L. \tag{1.4}$$

The presence of a prime in the quark doublet notation signifies that the spinor field is in the weak interaction eigenstates. The transformation from the eigenstate of mass to the eigenstate of weak interaction is achieved through the use of the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix for neutrinos and the Cabibbo-Kobayashi-Maskawa (CKM) matrix for quarks. The RH spinor field, denoted as $R(x) = \frac{1}{2}(1+\gamma^5)F(x)$, represents a weak isospin singlet labelled as $e_R^-, \mu_R^-, \tau_R^-, u_R, c_R, t_R, d_R, s_R, b_R$. The Standard Model does not include the presence of a right-handed neutrino or a left-handed anti-neutrino. It is important to be notice that the covariant derivative in Equation (1.3) behaves in a different way when acting on the left-handed (L) and right-handed (R) particles due to the V-A (vector minus axial vector) structure of the weak interaction. Additionally, the vector fields $W_\mu^a (a = 1, 2, 3)$ and B_μ are introduced to maintain the gauge symmetry. They are respectively further written

as

$$
\begin{aligned}
D_\mu L &= [\partial_\mu + ig'\frac{Y}{2}B_\mu + igT_a W_\mu^a]L \\
D_\mu R &= [\partial_\mu + ig'\frac{Y}{2}B_\mu]R \\
B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\
W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g f_{abc} W_\mu^b W_\nu^c.
\end{aligned} \tag{1.5}
$$

where the structure constants of the algebra of $SU(2)_L$, denoted as f_{abc}, play a crucial role. The weak hypercharge parameter Y, defined as $Y = 2(Q - I_W^3)$, combines the electromagnetic charge Q and the third component of weak isospin I_W^3. The generators $T_a (a = 1, 2, 3)$ represent the $SU(2)$ group and can be expressed using the Pauli matrices $\frac{\sigma_a}{2}$. The strengths of the couplings $U(1)$ and $SU(2)$ are respectively represented by g' and g in the forthcoming equations [2].

The gauge fields can be combined linearly to obtain the physical gauge bosons. The charged-current interaction straightforwardly determines the W bosons, $W_\mu^\pm$. On the other hand, the photon and the Z boson, A_μ and Z_μ, result from a linear combination due to the Higgs mechanism. They are respectively written as

$$
\begin{aligned}
W_\mu^+ &= \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2) \\
W_\mu^- &= \frac{1}{\sqrt{2}}(W_\mu^1 + iW_\mu^2) \\
A_\mu &= +B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \\
Z_\mu &= -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W.
\end{aligned} \tag{1.6}
$$

where, θ_W is the weak mixing angle and satisfies $\sin^2\theta_W \simeq 0.23$ and $g\sin\theta_W = g'\cos\theta_W$.

1.1.2 Electroweak spontaneous symmetry breaking

In the earlier section, the Lagrangians explicate massless gauge bosons and fermions. However, the inclusion of mass terms like $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$ and $-m_F\bar{F}F$ actually violates the local gauge symmetry. Within the framework of the Standard Model, the Higgs mechanism [11–16] spontaneously breaks the electroweak $SU(2)_L \times U(1)_Y$ local gauge symmetry, leading to the creation of particle mass and the existence of a massive scalar particle.

The Higgs field is selected by the SM to adopt the most cost-effective structure, identified as the Higgs doublet constituting a weak isospin pair with a weak hypercharge of $Y =$

[2] Throughout this chapter, we employ the symbol g to represent the quantity mentioned in some literature as g_W, where W signifies the weak interaction.

1.Comprised of four real scalar fields, the Higgs doublet is denoted as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.7}$$

Hence, the gauge invariant Lagrangian of the Higgs field may be referred to as $\mathscr{L}_{\text{Higgs}} = (D_\mu\phi)^\dagger(D_\mu\phi) - V(\phi)$, where the covariant derivative is written as $D_\mu = \partial_\mu + ig'\frac{Y}{2}B_\mu + igT_aW_\mu^a$ while the Higgs potential can be written as $V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$.

The Higgs potential has the coefficients that fulfill the condition of $\mu^2 < 0$ and $\lambda > 0$. This condition implies the existence of degenerate minima in the Higgs potential at $\phi^\dagger\phi = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda}$, which are also referred to as the vacuum.

In order for the photon to have no mass, it is crucial that the electromagnetic interaction's $U(1)$ gauge symmetry remains unbroken. Consequently, the determination of the Higgs doublet's vacuum expectation value (VEV), denoted as ϕ_0, is fixed at $\frac{1}{\sqrt{2}}\binom{0}{v}$. The determination of v is influenced by the Fermi coupling strength, denoted as G_F ($v = (\sqrt{2}G_F)^{-1/2} = 246$ GeV). Due to gauge invariance, the expansion from the vacuum can be described using the unitary gauge $\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$, with $h(x)$ denoting the Higgs field. The following expression is obtained by substituting expansion into the Higgs Lagrangian. As can be seen, its terms include the masses of the electroweak gauge fields, the coupling of the Higgs field to gauge boson fields, the massive scalar field h and the self-couplings of h. The resulting value for the mass of the physical scalar field h, also known as the Higgs boson, is $m_h = \sqrt{2\lambda}v$.

$$\begin{aligned} \mathscr{L}_{\text{Higgs}} = & \underbrace{\frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \lambda v^2 h^2}_{\text{massive scalar field } h} - \underbrace{\lambda v h^3 - \frac{\lambda}{4}h^4}_{h \text{ self-interaction}} \\ & + \underbrace{\frac{v^2}{8}g^2(W_\mu^1 W^{1,\mu} + W_\mu^2 iW^{2,\mu}) + \frac{v^2}{8}(gW_\mu^3 - g'B_\mu)(gW^{3,\mu} - g'B^\mu)}_{\text{terms that generate the mass of gauge bosons}} \\ & + \underbrace{\mathscr{L}_{\text{h, gauge}}}_{h\text{, gauge boson interaction}} . \end{aligned} \tag{1.8}$$

Thus, the determination of the λ parameter, signifying the self-interaction of the Higgs field, can be derived by assessing the mass of the Higgs boson, with λ equated to 0.13. Figure 1.1 illustrates the Higgs boson's self-interaction, with the coupling strengths labeled in proximity to the Feynman vertex.

The second row of Equation (1.8) provides the means to determine the masses of the physical gauge bosons. By directly examining the Lagrangian, one can extract the mass of the $W^{1,2}$ fields. This process is also applicable to the physical $W^\pm$ bosons. To obtain the masses of the photon and Z boson, one must diagonalize the mass matrix of the quadratic

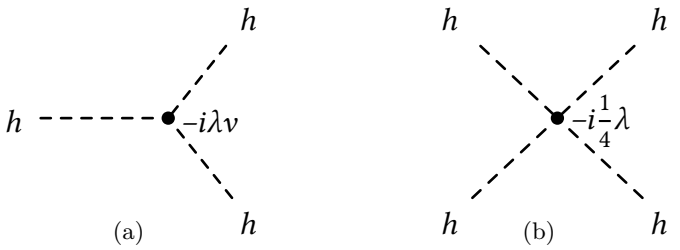


Figure 1.1: The self-interaction of the Higgs boson can be described using Feynman rules. There are two types of interactions: (a)trilinear and (b)quartic.

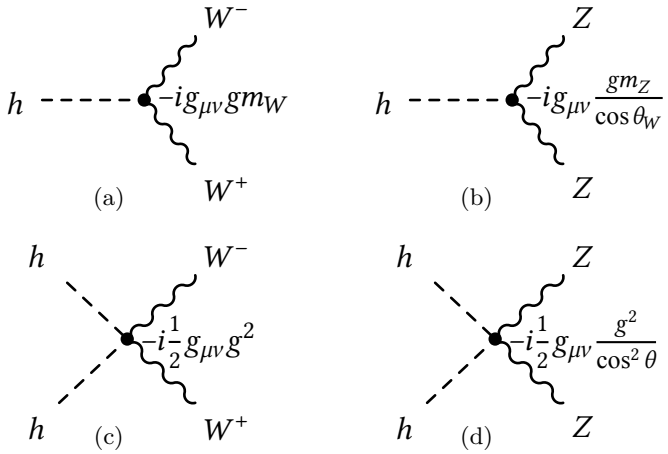


Figure 1.2: The interaction between the Higgs boson and $W^{\pm}/Z$ boson is governed by the Feynman rules, denoted as (a) hW^+W^-, (b) hZZ, (c) hhW^+W^-, (d) $hhZZ$.

term involving B and W^3. The physical gauge bosons have respective masses denoted as $m_{W^{\pm}} = \frac{1}{2}gv$, $m_\gamma = 0$ and $m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} = \frac{m_{W^{\pm}}}{\cos\theta_W}$. Within the Lagrangian of interactions between h and gauge fields $\mathscr{L}_{\mathrm{h,\ gauge}}$, there exist terms in the form of hVV and $hhVV$, where V represents a vector boson field. These couplings are visually depicted in Figure 1.2.

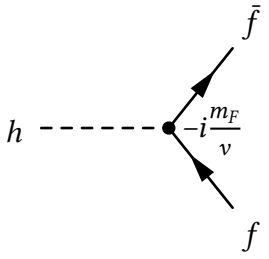


Figure 1.3: The coupling between the Higgs boson and fermion is described by the Feynman rule.

1.1.2.1 The mass of the fermions

The Higgs mechanism is responsible for generating the mass of fermions. In this process, the Higgs doublet is incorporated into the Dirac Lagrangian while maintaining the $SU(2)_L \times U(1)_Y$

gauge invariance. The Dirac Lagrangian is denoted as

$$\mathscr{L}_{\text{Fermion}} = \underbrace{-G_D(\bar{L}\phi R + \bar{R}\phi^{\dagger}L)}_{\text{for down type fermions}} - \underbrace{G_U(\bar{L}\phi_c R + \bar{R}\phi_c^{\dagger}L)}_{\text{for upper type fermions}} , \tag{1.9}$$

and includes the weak isospin doublet $L(x)$ and singlet $R(x)$ fermion fields, the weak hypercharge conjugate $\phi_c = -i\sigma_2\phi^*$ of the Higgs doublet ϕ, and the fermion flavor dependent coupling constants G_D and G_U for down type and upper type fermions respectively. Upon replacing the Higgs doublet's vacuum expansion, the Dirac Lagrangian for one fermion (F) transforms into

$$\mathscr{L}_{\text{Fermion}} = \underbrace{-\frac{G}{\sqrt{2}}v(\bar{F}_L F_R + \bar{F}_R F_L)}_{\text{fermion mass}} - \underbrace{\frac{G}{\sqrt{2}}h(\bar{F}_L F_R + \bar{F}_R F_L)}_{h\text{, fermion Yukawa coupling}} , \tag{1.10}$$

Within this Lagrangian, there exist a pair of components: the term for fermion mass, where $m_F = \frac{1}{\sqrt{2}}Gv$, and another term that describes the interaction between the Higgs boson and the fermion. The coupling is referred to as the Yukawa coupling, as illustrated in Figure 1.3. The determination of the Yukawa coupling constants G relies on the measured masses of the fermions.

1.2 Higgs boson and its pair production

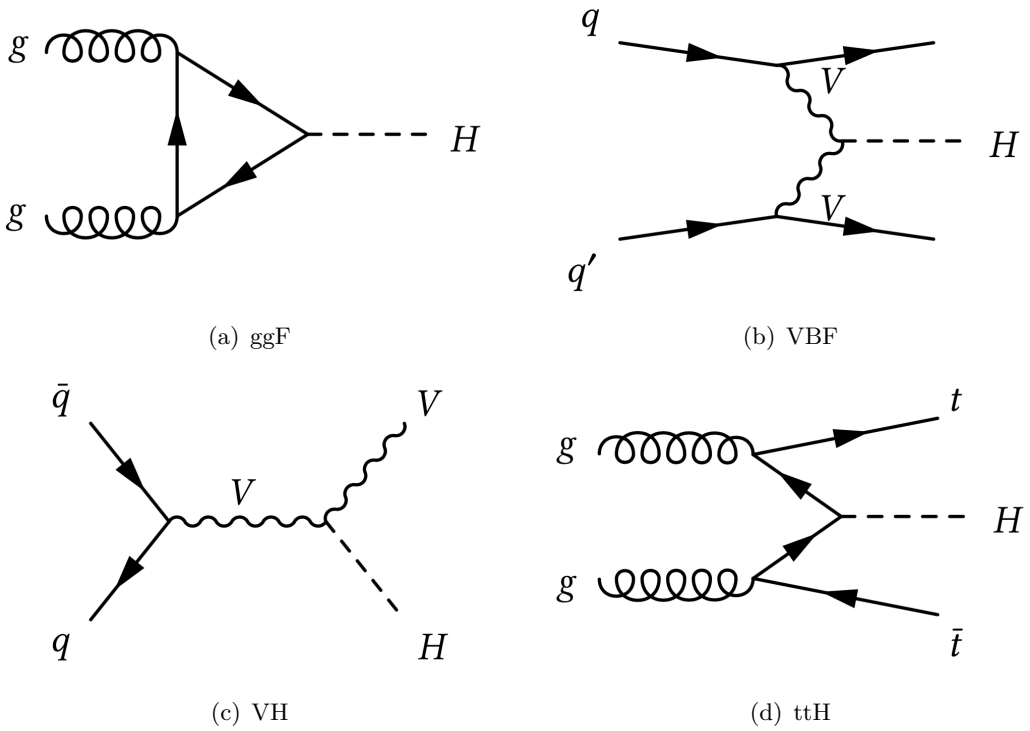
1.2.1 Higgs boson properties

In the year 2012, a resonance particle with a mass of around 125 GeV was detected by the ATLAS [1] and CMS [2] experiments. This particular particle, known as the SM Higgs boson (H) [3], is unique as it is the only scalar particle ever observed. Its properties align remarkably well with those predicted by the electroweak symmetry breaking mechanism for the Higgs boson. The experimental value for the mass of the Higgs boson is $m_H = 125.10 \pm 0.14$ GeV, and its spin-parity is $J^P = 0^+$ [17].

The Higgs boson is primarily generated at the LHC through several processes. These include gluon-gluon fusion via a loop of t/b-quark (ggF), vector boson fusion (VBF), vector boson associated process (VH, which includes WH and ZH), and top-quark pair associated process (ttH). Figure 1.4 provides a summary of the Leading Order (LO) Feynman diagrams for these production modes. Theoretical calculations have been made to predict their cross sections at a centre-of-mass (COM) energy of $\sqrt{s} = 13$ TeV and a Higgs boson mass of 125 GeV, and they are respectively 49 pb (ggF), 3.8 pb (VBF), 1.4 pb (WH), 0.88 pb (ZH) and

[3] To provide further clarification on the notation, it should be noted that in the previous chapter, the lowercase h was used to represent both the Higgs field and the physical particle known as the Higgs boson. However, starting from this chapter onwards, the capital H will exclusively denote the physical Higgs boson.

0.51 pb (ttH). The cross sections for these processes are determined using advanced theoretical predictions in the high order perturbation calculation of QCD and EW theory [18].



(a) ggF (b) VBF

(c) VH (d) ttH

Figure 1.4: The Feynman diagrams at the leading order for the main production modes of the Higgs boson through proton-proton collisions at the LHC.

According to Section 1.1.2, the SM Higgs mechanism establishes the connections between the Higgs boson and vector bosons as well as fermions. These connections are determined by the masses of the particles, with the coupling strengths being proportional to their masses. This information allows us to make predictions about the decay mechanism of the Higgs boson. It is important to note that in the process $H \longrightarrow VV$, one of the vector bosons must be off-shell due to the mass threshold. Additionally, the Higgs boson just effectively couples to massless photons and gluons through a loop diagram. The predicted theoretical branching ratios for the primary decay channels of the Higgs boson with an assumed mass of 125 GeV, are listed below:

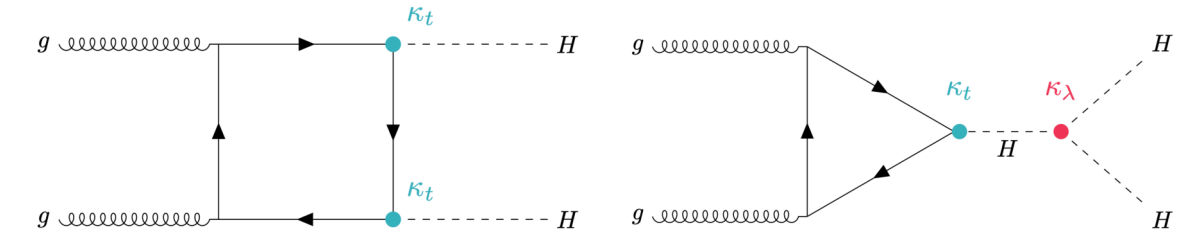
- $H \longrightarrow b\bar{b}$ ($58.2^{+1.2}_{-1.3}$ %)
- $H \longrightarrow W^+W^-$ (21.4 $\pm$1.5 %)
- $H \longrightarrow \tau^+\tau^-$ (6.27 $\pm$1.6 %)
- $H \longrightarrow ZZ$ (2.62 $\pm$1.5 %)
- $H \longrightarrow \gamma\gamma$ (0.227 $\pm$2.1 %)

The Higgs boson's mass is a parameter in the SM that can vary. This mass also determines the value of the Higgs boson self-coupling constant, denoted as λ. The search for Higgs boson pair (HH) production is driven by the desire to directly measure λ, as changing

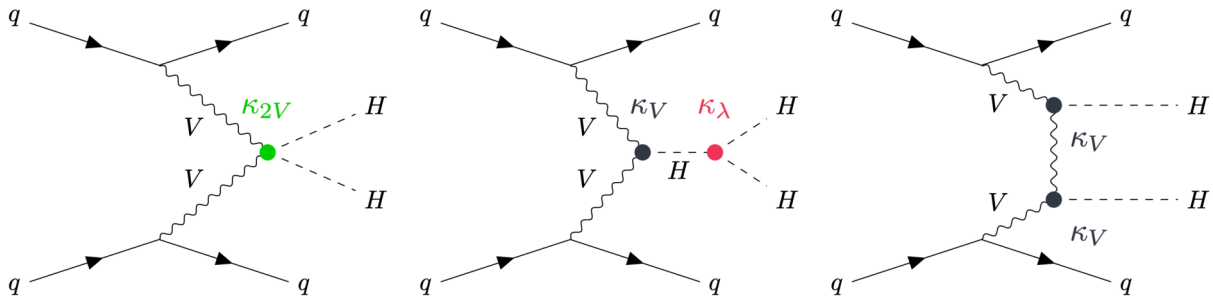
its value can significantly impact the production cross section of HH (denoted as σ_{HH}). The relationship between σ_{HH} and λ is explained in Section 1.2.2. Typically, the measurement is conventionally performed on the coupling modifier $\kappa_\lambda = \frac{\lambda}{\lambda_{SM}}$, where $\lambda_{SM} \simeq 0.13$ represents the SM value of λ. In the subsequent subsection, a comprehensive explanation of the production and decay mechanism of Higgs boson pair is provided.

1.2.2 Higgs boson pair production

The production of a pair of Higgs boson is similar to the production of a single Higgs boson, and it is also with the dominant modes being gluon-gluon fusion and vector boson fusion. Figure 1.5 provides a summary of the LO Feynman diagrams.



(a) ggF



(b) VBF

Figure 1.5: The leading order Feynman diagrams for the two main production modes of Higgs boson pair via proton-proton collision at the LHC, (a) ggF and (b) VBF.

The Next-Next-to LO (NNLO) precision [4] [18] calculation of the ggF cross section, using the FTapprox method [19], represents a state-of-the-art theoretical prediction. The uncertainty associated with this calculation is provided by Ref. [20]. In the FTApprox method, real radiation corrections are computed with a finite top mass, while the virtual loop corrections are based on Higgs Effective Field Theory (HEFT). The uncertainties arise from various factors, including the parton distribution function and α_S of the proton, which are based on PDF4LHCnnlomc, the scale of re-normalization and factorization, as well as the choice of re-normalization scheme and scale in the calculation of the top-quark mass.

[4] In this particular section, the level of accuracy for the calculations is referred to as QCD precision unless otherwise specified. On the other hand, the electroweak calculation consistently maintains a precision level of NLO.

Moving on to VBF, the cross section is determined at Next-Next-Next-to LO (N3LO). [21] The cross section of the ggF single Higgs is $\sigma_{\text{ggF}} = 31.05\ \text{fb} \pm 3.0\%\ ^{+2.2\%}_{-5.0\%}\ ^{+4\%}_{-18\%}$, where the three types of uncertainties are respectively from PDF and α_S, Scale, and m_{top}. The cross section of the VBF single Higgs is $\sigma_{\text{VBF}} = 1.726\ \text{fb} \pm 2.1\%\ ^{+0.03\%}_{-0.04\%}$, where the two types uncertainties are respectively from PDF and α_S, and Scale.

The production of a pair of Higgs bosons has the cross section that is approximately one thousandth of the cross section for a single Higgs boson. This is due to the destructive interference between the 'triangle' and 'box' diagrams depicted in Figure 1.5(a). The reduced cross section poses a significant challenge in the search for these events. However, it is possible to modify the production cross section by adjusting the self-coupling constant of the Higgs boson. Figure 1.6 illustrates the relationship between the Higgs production cross section and the κ_λ variable.

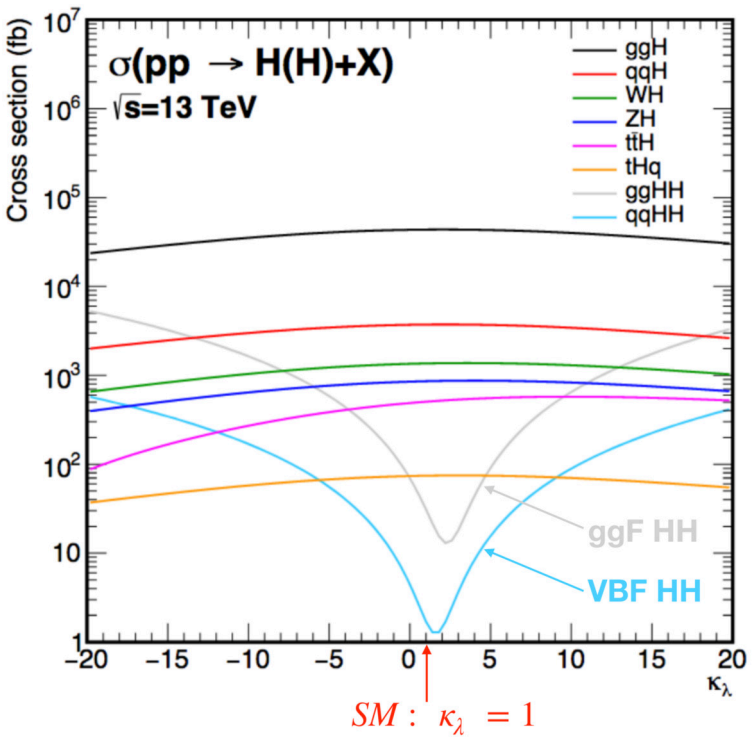


Figure 1.6: The production cross section of single Higgs and Higgs boson pair as a function of κ_λ [22]. The grey and cyan curves are respectively for ggF and VBF Higgs boson pair. The red vertical line represents the SM case $\kappa_\lambda = 1$.

The simulated HH samples used in Chapter 3 are normalized using the cross sections. These samples are generated by POWHEG BOX V2 [23] with full top-quark mass (FT) at Nextto LO (NLO) precision. In the FT method, both real and virtual corrections utilize the finite mass of the topquark. The reason for using the FT method to generate the HH samples is that it provides more accurate predictions of Higgs boson kinematics [24]. Ref. [25] presents a comparison between the kinematics of the HH system in this sample and the FTApprox-based sample generated using MADGRAPH5_aMC@NLO [26, 27]. These

comparisons are crucial in determining the ggF sample generators for HH physics analysis in ATLAS.

The Higgs boson pair decay is actually the result of combining the decays of each individual Higgs boson. Hence, the branching ratios of the primary decay channels of HH can be determined by considering the branching ratios of single Higgs decay channels. The following list presents these branching ratios:

- $HH \longrightarrow b\bar{b}b\bar{b}$ (34%)
- $HH \longrightarrow b\bar{b}W^+W^-$ (25%)
- $HH \longrightarrow W^+W^-W^+W^-$ (4.6%)
- $HH \longrightarrow b\bar{b}\tau^+\tau^-$ (7.3%)
- $HH \longrightarrow b\bar{b}\gamma\gamma$ (0.26%)

1.2.3 Higgs potential properties

The measurement of the self-coupling of the Higgs boson holds significant significance in comprehending the mechanism of electroweak symmetry breaking. Furthermore, the self-coupling provides insights into the overall characteristics of the Higgs potential, which is essential for addressing profound physics inquiries. This leads us to briefly introduce two intriguing subjects (stability of vacuum and electroweak baryongenesis) that are closely connected to the global properties of the Higgs potential. The primary objective in the search for Higgs pair production is to investigate any phenomena that could potentially influence the overall characteristics of the Higgs potential.

The phase of the Standard Model is determined by the shape of the Higgs potential. At the electroweak scale, the potential at the tree level provides a satisfactory approximation. Nevertheless, when the Higgs field h increases in energy, it becomes necessary to consider quantum corrections, assuming that the Standard Model remains valid up to the Planck scale Λ_{Pl}, which is about 10^{19} GeV. The SM parameters are computed with NNLO precision at the EW scale and then extrapolated to the Planck scale using Re-normalization Group Equations (RGE) inRef. [28]. The Higgs potential undergoes significant changes in shape due to the evolution of these parameters from low to high energy scales. It has been observed that the Higgs quartic coupling constant and the top-quark Yukawa coupling constant are highly sensitive to the energy scale, as shown in Figure 1 of Ref. [29]. The current measurements of the masses of the Higgs boson and the top-quark indicate that the SM phase is close to the boundary between a stable and a metastable phase, and it is compatible with a metastable vacuum [28]. Consequently, there exists a possibility for the SM to transit to the true vacuum through the quantum tunnelling effect. This has significant implications from a cosmological standpoint, as the nature of our universe would be completely different if the tunnelling rate is sufficiently high to induce a phase transition. Ref. [30] discusses the

decay rate of our EW vacuum, which is found to be extremely small. Therefore, we are safe unless the decay rate is significantly altered by a new physics mechanism.

The finite temperature quantum field theory allows for the simulation of the Higgs potential's evolution during the thermal history of the early universe. This simulation sheds light on the dynamics of the electroweak phase transition (EWPT), where the electroweak theory undergoes a transformation from a symmetric phase to a symmetry breaking phase. Through this evolution, the nature of the EWPT, whether it is a first-order or second-order phase transition, is unveiled.

The electroweak baryogenesis (EWBG) [31] necessitates a first-order electroweak phase transition (EWPT), and it is a highly promising theory for explaining the observed matter-antimatter asymmetry in the universe. Figure 1.7 provides a schematic representation of how the shape of the Higgs potential changes as the universe cools, assuming a first-order phase transition for the EWPT. Initially, at very high temperatures, only a phase with zero vacuum expectation value (VEV) exists. However, once the temperature reaches a critical value, denoted as T_c, a degenerate vacuum with the VEV of being ϕ_c is formed. As the universe continues to cool towards $T \longrightarrow 0$, the symmetric phase undergoes a tunneling process to transit into the symmetry-breaking phase. Ultimately, the minimum field value converges to the electroweak vacuum v.

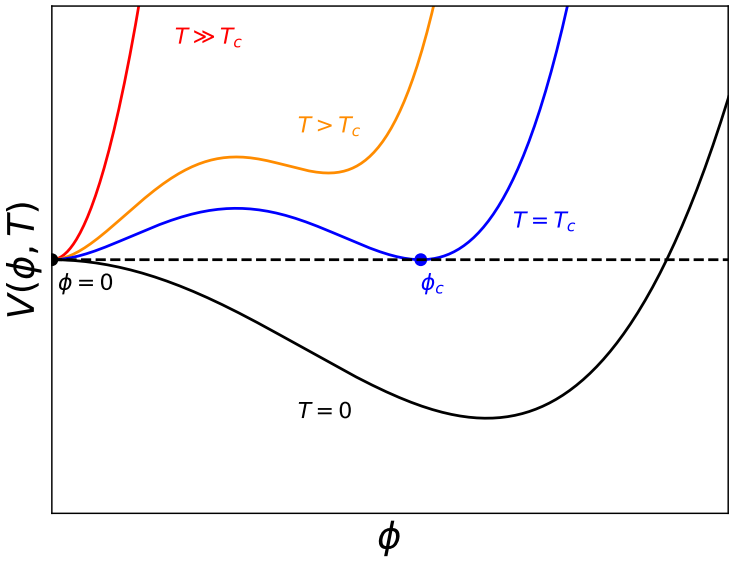


Figure 1.7: The Higgs potential that varies depending on the thermal conditions in the universe and assuming the first-order electroweak phase transition, is depicted in a schematic illustration. It is important to note that the scale represented in the figure does not correspond to the actual scale of h or $V(h)$.

Moreover, it is crucial for the phase transition to be robust, known as a strong first-order phase transition, in order to prevent the baryon numbers from being erased by sphalerons. The condition that must be met is $\frac{\phi_c}{T_c} \gtrsim 1$. However, the observations on the Higgs boson suggest that the EWPT within the Standard Model is a second-order phase transition,

also referred to as a crossover [32, 33], where the VEV smoothly transitions from 0 to v. Consequently, new physics beyond the Standard Model is necessary to achieve the desired strong first-order phase transition. For example, according to Ref. [34], it is predicted that if the strength of the self-coupling of the SM Higgs boson is enhanced by more than 50% (meaning $\kappa_\lambda \gtrsim 1.5$), then the EWPT can transform into a strong first-order phase transition.

1.2.4 Beyond-the-SM physics and the HH production

The one of the main primary objectives at LHC is to investigate the global characteristics of the Higgs potential. It focuses on detecting the creation of Higgs boson pairs as a means to achieve this goal. This is because alterations to the Higgs potential can impact the cross section of Higgs boson pair production. If deviations from the standard model prediction are observed in the cross section of Higgs boson pair production, it will revolutionize our comprehension of the Higgs potential.

This thesis aims to investigate the non-resonant HH production mode. The search is carried out in a model-independent manner, allowing for sensitivity to any potential new HH production signals. The primary signal mode used as a baseline in this search is the SM ggF+VBF HH production, as depicted in Figure 1.5.

Different BSM models anticipate an increase in the rate of Higgs pair production through non-resonant processes. This can occur either due to the involvement of new particles in the loop diagram during production or through anomalous couplings such as non-SM Higgs boson self-coupling or non-SM top-quark Yukawa coupling [35, 36]. The production of HH via ggF and VBF modes is fundamentally influenced by the coupling modifiers κ_λ, κ_t, κ_V and κ_{2V}. The ggF production mode, due to its higher cross-section, provides the most effective means to probe κ_λ. On the other hand, the VBF topology exhibits a distinct sensitivity to κ_{2V} as the ggF process does not incorporate the $HHVV$ interaction.

The computational cost and time required for a thorough simulation of HH samples using a detailed grid in the (κ_λ,κ_{2V}) plane are significant. To address this limitation, a reduced number of MC simulation samples are generated for specific coupling values, and a technique for combining these samples is utilized to represent the signal hypothesis across the entire coupling parameter space. The technique of combining multiple samples in a manner that encompasses the complete range of coupling constants within the parameter space relies on exploiting the fundamental principles of the differential cross-section formula.

Chapter 2

The ATLAS Experiment

This chapter presents an overview of the ATLAS experiment. Firstly, it briefly introduces the LHC, followed by a description of the ATLAS detector including the inner detector, calorimeters, muon spectrometer, trigger and data acquisition. At the end, the definition of physics objects is presented.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest particle physics experimental platform in the world composed of state-of-the-art hadron accelerators and collider [37–39]. The primary goal of the platform is to accelerate protons to nearly the speed of light and collide them at some specific collision points, and heavy-ions can also be run in this machine. The maximum centre-of-mass energy is designed to be $\sqrt{s} = 14$ TeV while its peak luminosity can reach up to $\mathscr{L} = 10^{34}$ cm^{-2}s^{-1}. There are two transfer tunnels with the length of about 2.5 kilometers dedicated for the connection between CERN accelerator complex and the LHC ring that is 27 kilometers in circumference and installed about 100 meters (to be specific, from 45 to 170 meters) under the ground, also for the injection of hadron beams towards the LHC ring.

The hadron beams are structured in the form of consecutive bunches and synchronized with the 40 MHz LHC clock. Spatially, in terms of proton runs, there are 2808 bunches running in the LHC ring a the same time and each of them is made of about 10^8 protons. Temporally, the time interval between two adjacent bunches is typically 25 nanoseconds, so that the bunches also cross every 25 nanoseconds, which is exactly one cycle of the LHC clock.

Once the hadron beams are injected into the LHC ring, they will go through the processes of capture and acceleration, and finally be stored in the superconducting radio frequency cavities. The hadron beams are actually handled in two separate cavities and thus the resulting two types of beams run in the opposite direction at the relativistic speed. The opposite beams are precisely controlled to interact with each other at four main collision points and some smaller ones along the LHC ring, and corresponding experiments are es-

tablished around those points. The four main LHC experiments include the generalpurpose ATLAS [40] and CMS [41] experiments designed to search for Higgs boson and supersymmetry, and explore SM and diverse exotic physics, the LHCb [42] experiment for B-physics research, and the ALICE [43] experiment for probing quarkgluon plasma by virtue of the product from heavy-ion collisions.

Figure 2.1 shows the schematic of the CERN accelerator complex where the step-up facility structure is adopted to realize the final hadron beams with desired quality. In terms of proton beams injection, they are firstly fed into the Linac that is connected to the Booster. After acceleration of the Booster, they are transferred to the CERN Proton Synchrotron (PS) followed by the Super Proton Synchrotron (SPS), after which, they are finally injected to the LHC ring.

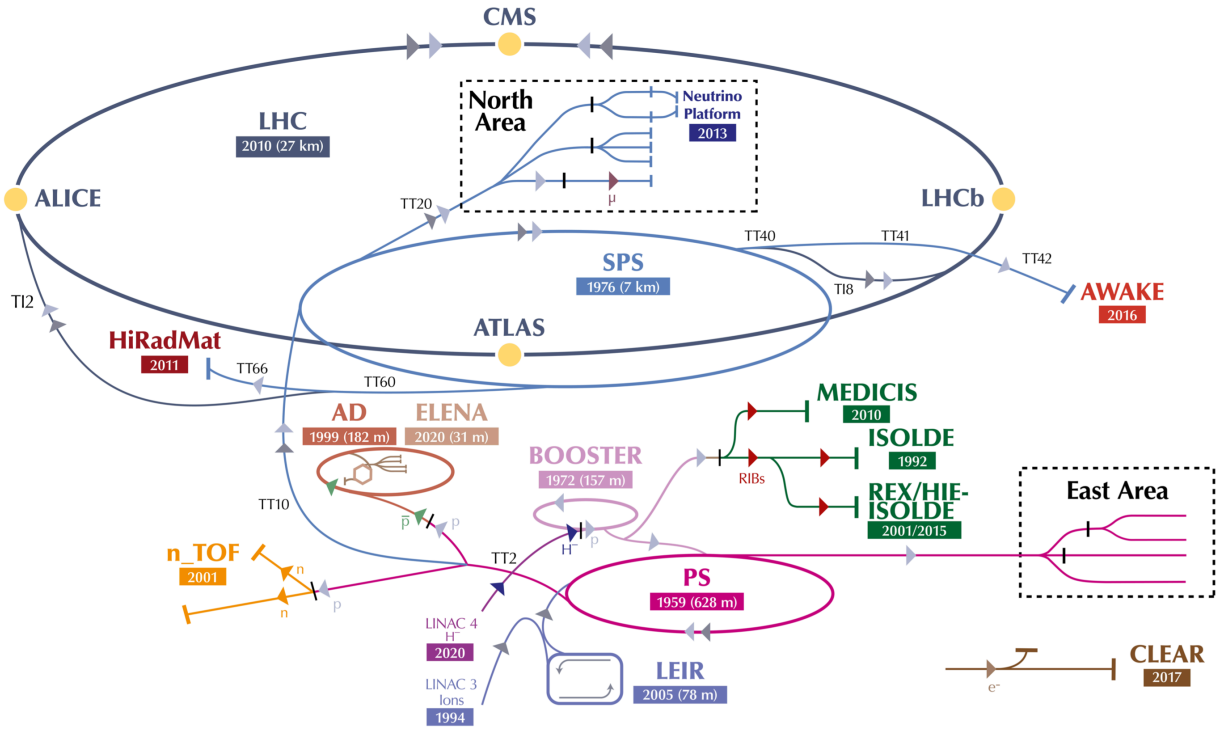
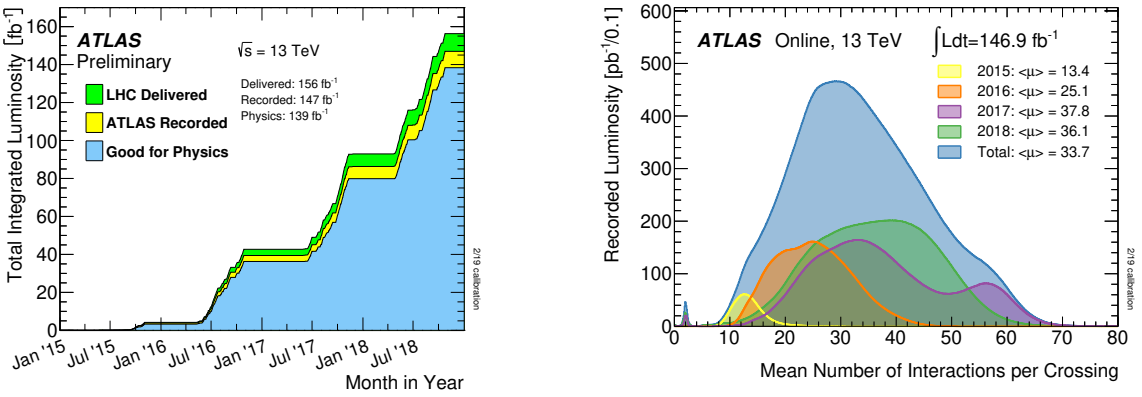


Figure 2.1: The schematic of the CERN accelerator complex [44].

It's scheduled that the LHC operates in three periods, the Run-1 in the years of 2011 and 2012 during which the Higgs boson was discovered, the Run-2 lasting from 2015 to 2018, and the Run-3 lasting from 2022 to 2025. For the Run-1 period, the target centre-of-mass energy was set to be $\sqrt{s}$ = 7 and 8 TeV, which led to the collection of p-p collision dataset with the integrated luminosity of L =28.3 fb^{-1} in the ATLAS detector. For the Run-2 period, the target centre-of-mass energy was upgraded to 13 TeV and the instantaneous luminosity was also increased, especially for the operations in the years of 2017 and 2018. Finally, it gave rise to the delivery of L =156 fb^{-1} p-p collision data in the ATLAS detector during Run-2. It should be noted that, among the 156 fb^{-1}, only 139 fb^{-1} was verified and used for physics analysis. Figure 2.2 shows the evolution of total integrated luminosity with time and the

mean number of per-bunch-crossing interactions $\langle\mu\rangle$ (also called pile-up) for the Run-2 data taking. As can be seen, the instantaneous luminosity in the years of 2017 and 2018 with corresponding $\langle\mu\rangle$ of respectively 37.8 and 36.1, a is significantly increased so that about 75% of LHC Run-2 dataset was taken in these two years.



(a) Total integrated luminosity

(b) Pile-up

Figure 2.2: Plots of the total integrated luminosity as a function of time and the pile-up in Run-2 data taking.

The LHC is currently in operation with the centre-of-mass energy of $\sqrt{s}$ = 13.6 TeV for the Run-3 data taking. Its instantaneous luminosity is set to be $\mathscr{L} = 2 \times 10^{34}$ cm^{-2}s^{-1}, and it's expected that 300 fb^{-1} more collision data will be collected at the end of Run-3. Right after LHC Run-3, the High-Luminosity LHC (HL-LHC) is scheduled to take more data with a much higher instantaneous luminosity of 5×10^{34} cm^{-2}s^{-1}. Section 4.2 describes the HL-LHC in details. To deal with the more harsher operational environment, the LHC machine and its experiments will be upgraded. Ultimately, the HL-LHC is expected to deliver up to 3000 fb^{-1} dataset to each of the main experiments.

2.2 The ATLAS detector

The ATLAS detector [40] is the largest particle physics detector human has ever built. It is installed at one of the collision points along the LHC ring, generally speaking, taking the shape of a cylinder with the length of 44 meters and the diameter of 25 meters. Surprisingly, it has the weight of about 7000 tons. Figure 2.3 gives a general depiction of the ATLAS detector. As can be seen, from inside to outside, the inner detector, the calorimeters (including electromagnetic and hadronic parts), and the muon spectrometer are sequentially arranged.

When passing through the ATLAS detector, the outgoing particles interact with each of the sub-detectors along the way and leave hits and energy deposition in the sensitive cells of the sub-detectors. Based on the acquired information from them, the ATLAS detector

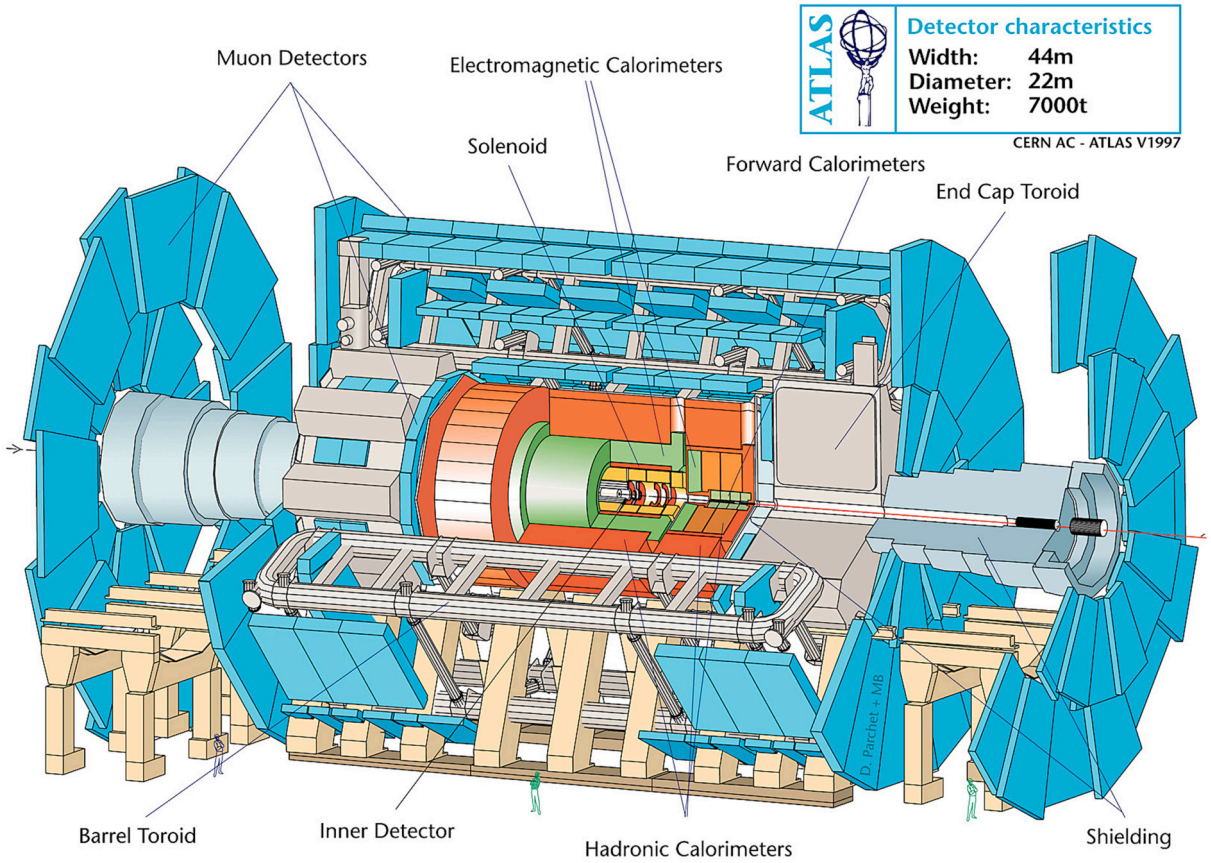


Figure 2.3: Demonstration of the ATLAS detector [45].

can provide excellent measurements on those particles generated from the collision point, such as their momentum, energy and particle identification. For the particle types that the detector measures, they can be any particles that interact with the detector materials and leave visible products. For example, the target particles can be electron, photons, muon, pion, neutron and proton.

In general, there are two magnet systems in the ATLAS detector, which are the solenoid system and the toroid system. The former is installed around the inner detector and provides an axial magnetic field of up to 2 teslas for deflecting the trajectory of the charged particles inside the inner detector while the latter is installed in both the barrel area (barrel toroid) and end-cap area (end-cap toroid) around the muon spectrometer and provides the magnetic fields of 0.5 and 1 tesla for bending muon trajectories.

The coordinate system of the ATLAS experiment is established with the interaction point (IP) as the origin. The three axises is defined right-handedly, which means the x-axis points towards the centre of the LHC ring; the y-axis points upwards; and the z-axis is in line with the tangential direction of the beamline. In the x-y plane, the azimuthal angle ϕ is defined while the polar angle θ is defined in the y-z plane. Usually, instead of using the θ, the pseudorapidity $\eta = -\ln(\tan\frac{\theta}{2})$ is defined to describe the angle of a particle relative to the beam axis in the aspect of Lorentz boost. Therefore, the direction of y-axis has the

pseudorapidity value of zero while the direction of z-axis has the pseudorapidity value of positive infinity. In addition, in order to measure the angular distance between two four-vectors in $\phi - \eta$ space, an angular quantity is defined with the formula $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, which is Lorentz invariant if the involved particles are massless.

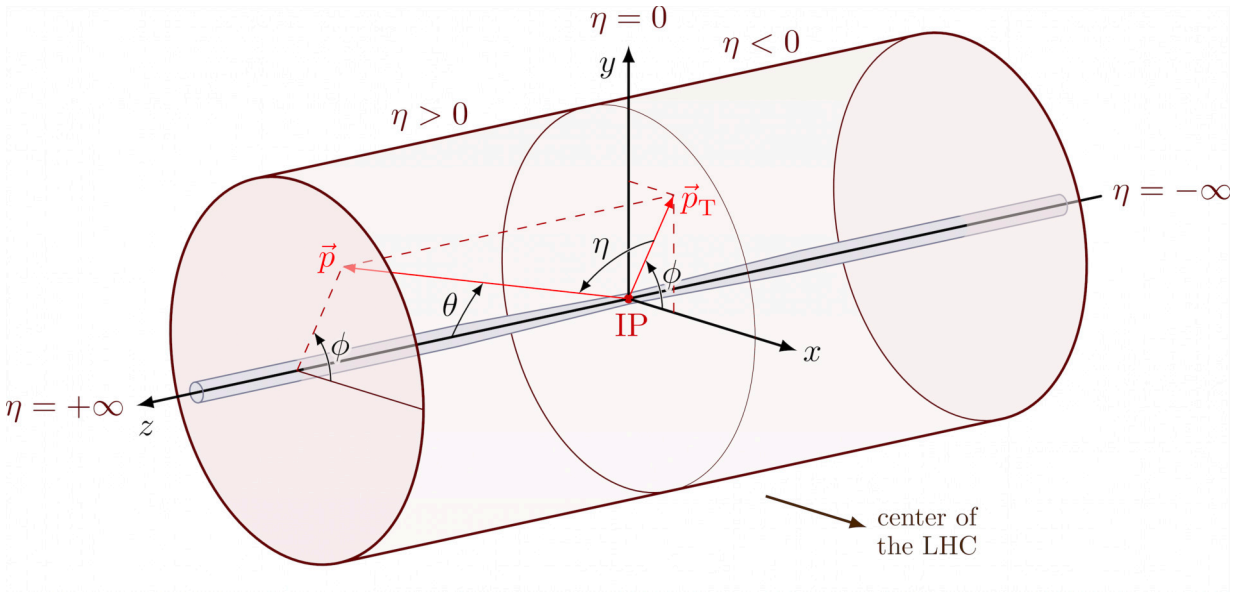


Figure 2.4: Coordinate system of the ATLAS experiment [46].

2.2.1 Inner Detector

The Inner Detector is located in the innermost part the ATLAS detector. The outgoing particles from interaction point pass through sensitive cells (pixels, strips or straw tubes) of the detector and cause ionization process, which leads to an electrical pulse in each cell along its trajectory and leaves a series of hits. With the help of dedicated algorithm, those hits can be used to deduce the information of the leaving charged particles, such as their vertexes and tracks, with which the impact parameters and momentum can be calculated.

The inner detector can be further divided into three parts, which are arranged from inside to outside with the order of the pixel detector, SCT (silicon microstrip tracker) and TRT (Transition Radiation Tracker).

Figure 2.5 shows the structure of the ATLAS inner detector. The detector is quite compact and just extends to the radius of about one meter from the beamline. The pixel detector is arranged with three layers around the beamline, extending to the radius of about $12.3cm$. During the LS1 (long shutdown 1) of the LHC, in order to enhance the vertex reconstruction, an Insertable B-Layer (IBL) [47] was inserted in the barrel region of the pixel detector. In the end-cap regions of the pixel detector, there are also three disks that are laid out perpendicular to the beamline. Both the pixel detector and SCT have a good coverage of $|\eta| < 2.5$. The SCT is also arranged both in the barrel region with eight strip

layers and in the end-cap region with nine disks, similar to the layout of the inner detector. The TRT is consisted of straw tubes placed in parallel with beamline. Because of this nature, the z positions of a track can not be acquired with the TRT. Its feature of detecting transition radiations can significantly benefit the identification of electrons. The coverage of the TRT is a bit smaller, $|\eta| < 2.0$.

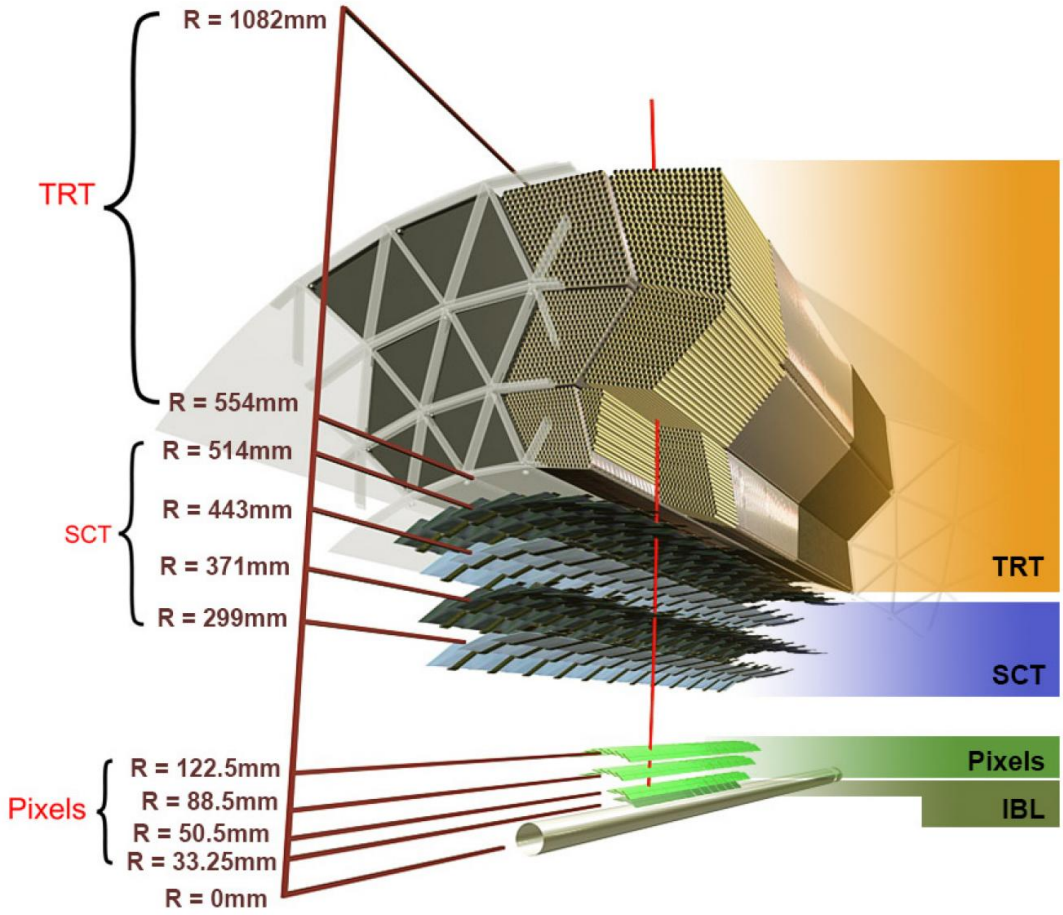


Figure 2.5: Structure of the ATLAS inner detector [48].

2.2.2 Calorimeters

Surrounding the inner detector, the calorimeter is installed in both the barrel region and end-cap region. The mechanism of the calorimeter is to stop the charged particles like electrons, photons and hadrons, and measure the ensuing showers of them left in the sampling materials. The showers are actually the collection of secondary particles caused by the cascaded processes originated from the primary incident charged particle. There are two kinds of showers in the calorimeter, which are electromagnetic shower and hadronic shower. The former is caused by electrons or photons while the latter is usually caused by incident charged hadrons. The sensitive materials absorb the deposited energy of the showers and

digitize it into electrical signals that are finally acquired by the DAQ system.

In terms of the electromagnetic calorimeter (ECal), it is located outside the inner detector and close to the solenoid system. The operation material of ECal is made of liquid-argon (LAr) with lead as absorber and kapton as electrodes plates. Figure 2.6 shows the structure of ECal. There are in total four layers in ECal, which are three active layers (EM1, EM2, EM3) and one presampler layer. Each of the three active layers has different designs. The segmentation of the layers actually depends on the pseudorapidity η. Taking $\eta = 0$ as an example, the thickness and granularity ($\Delta\eta \times \Delta\phi$) of the layers are specifically designed like this. The presampler layer with the thickness of $11\ mm$ and granularity of 0.025×0.1 is used to correct the energy loss ahead of the ECal; the EM1 has $4.2X_0$ thickness and 0.003×0.1 granularity to discriminate neutral pion and photon; the EM2 has $16X_0$ thickness and 0.025×0.025 granularity for collection of the majority of the energy deposition; the EM3 is designed with the thickness of $2.0X_0$ and the granularity of 0.05×0.025 to collect the energy in the tail of the shower. The X_0 is radiation length and has the value of $X_0 = 14.2$ cm in this case [49, 50]. The selections of their thickness and granularity are finally determined by dedicated optimization studies such that they are long enough to stop the showers and completely absorb deposited energy, and also their granularities are fine enough to precisely measure the energy of electrons and photons. More specifically, in order to distinguish photon from the photon of neutral pion decay ($\pi^0 \longrightarrow \gamma\gamma$), finer granularity is considered in the EM1. In addition, to better discriminate close photons, the separation in η is even finer than that of two photons originating from π^0 decay.

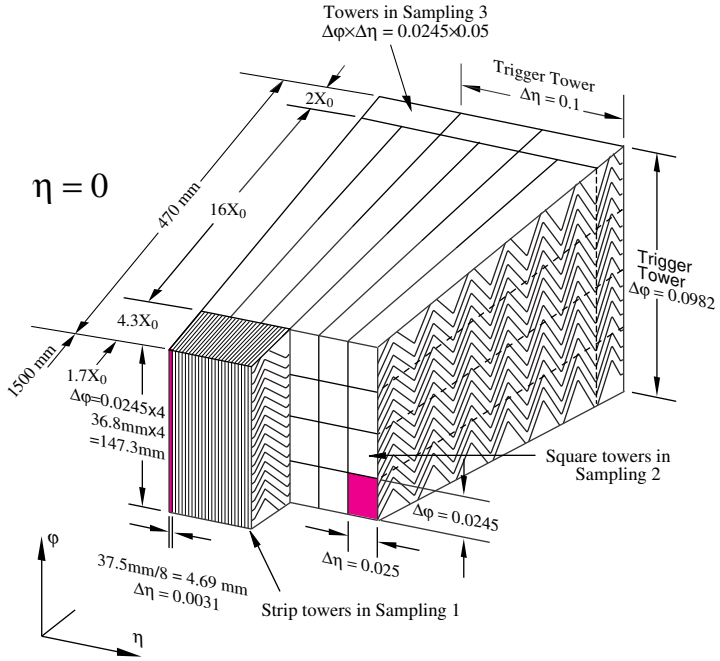


Figure 2.6: The accordion structure of the EM calorimeter [50].

The good coverage of ECal comes from the barrel coverage of $0 < |\eta| < 1.475$ and end-cap

coverage of $1.375 < |\eta| < 3.2$, where the end-cap coverage further comes from two wheels, one covers $1.375 < |\eta| < 2.5$ while the other covers $2.5 < |\eta| < 3.2$. Due to the inevitable gap between the barrel and end-cap in the detector arrangement, the ensuing crack region of $1.37 < |\eta| < 1.52$ is not installed with active layers so that the ECal can not provide measurement on showers over there.

The hadronic calorimeter (HCal) is also installed in the barrel region and end-cap region right after the ECal. Figure 2.7 shows the installation position of the HCal. In the barrel region, the scintillator-tile calorimeter is installed, covering the region of $|\eta| < 1.0$ and $0.8 < |\eta| < 1.7$. It adopts the plastic scintillators as the active material while the steel as its absorber. After particles hit the scintillators, lights are generated and immediately go to the optical fibers that can shift the wavelength of the lights. Further, the wavelength-shifted lights are detected by the photomultiplier tubes whose output is proportional to the energy deposition of the hadronic showers. In the end-cap region, there are two parts, the hadronic end-cap calorimeter (HEC) covering the region of $1.5 < |\eta| < 3.2$, and the forward calorimeter (FCal) covering the region of up to $|\eta| = 4.9$. Both of them uses liquid-argon as their active material. The HEC actually adopts the same technique with ECal, but the slight difference is that its absorber is chosen as copper. Both electromagnetic and hadronic showers can be measured by the FCal, also as the last parts before muon spectrometer, the FCal is designed to be dense enough so that it can prevent the radiation background from contaminating the muon measurement.

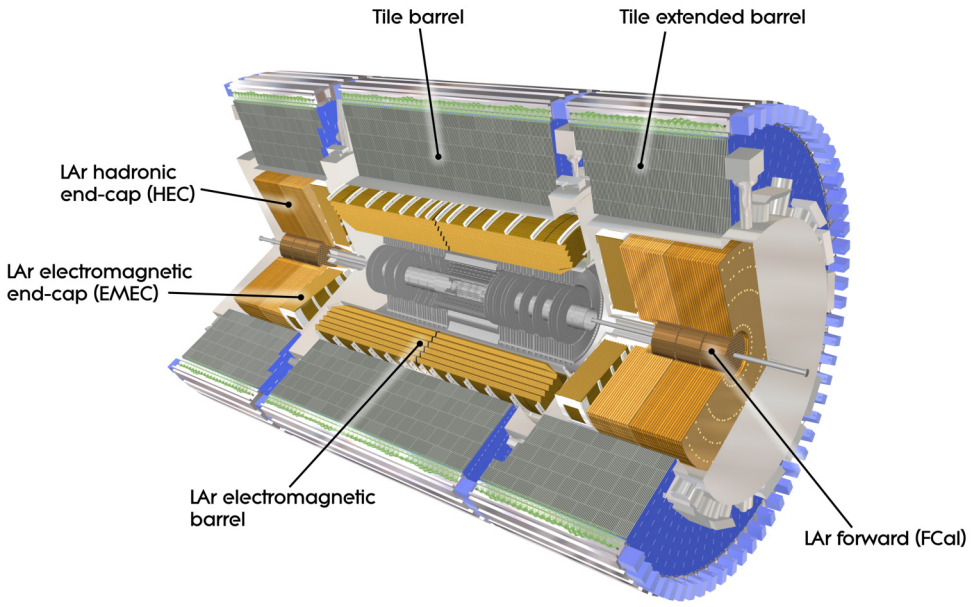


Figure 2.7: The position of the hadronic calorimeter [51].

2.2.3 Muon spectrometer

The muon spectrometer is the last sub-detector of the ATLAS detector and located in the outermost area. Figure 2.8 shows the installation position of the muon spectrometer in the ATLAS detector. With the magnetic field provided by the toroid magnetic system, the muon track can be deflected and thus its momentum can be accurately measured by the muon spectrometer. Similarly, the spectrometer is also distributed in both the barrel region and end-cap region. Respectively, the Monitored Drift Tube (MDT) and Cathode Strip Chamber (CSC) are installed in those two regions. The muon spectrometer has its own triggering chambers. The Resistive Plate Chamber (RPC) is mainly used to trigger the detector in the barrel region while the Thin Gap Chamber (TGC) is used to trigger the detector in the end-cap region.

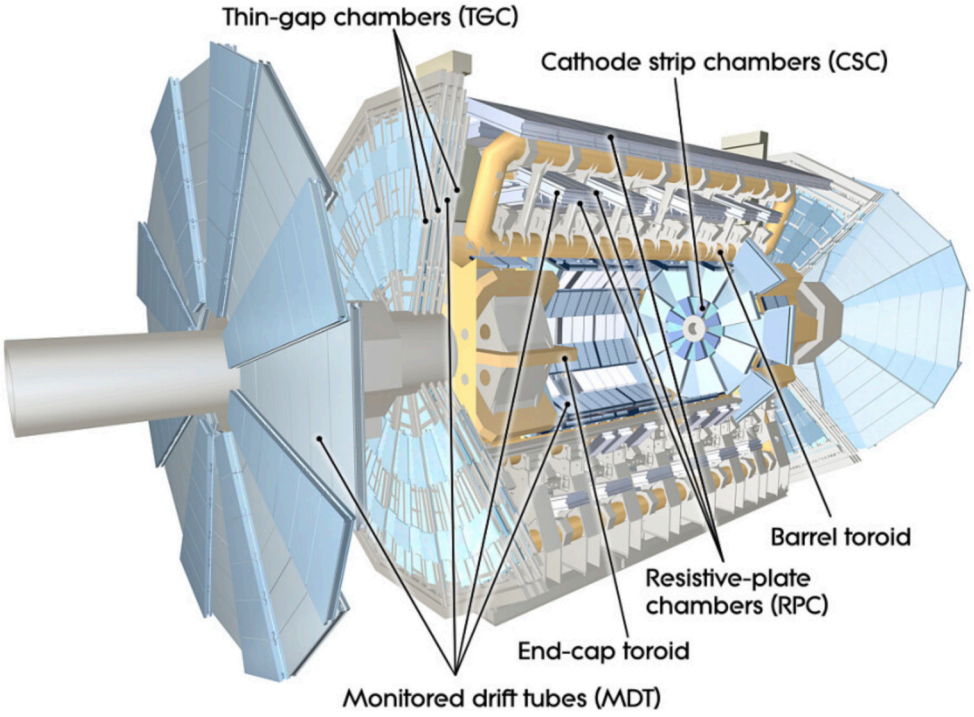


Figure 2.8: The position of the muon spectrometer [52].

2.2.4 Trigger and data acquisition

The LHC collides the proton bunches under the clock of 40 MHz. It is impossible to record all the events from the bunch crossing since that requires tremendous amount of storage space. The solution is to just trigger the events of interest with a lower frequency, then they can be easily handled by the data acquisition system. With the two-level trigger system [53, 54] deployed in the ATLAS detector, the event rate can be greatly reduced to about 1 kHz. In addition, a trigger and data acquisition (TDAQ) system is designed to process the event stream offline.

The two-level trigger system is consisted of the hardware-based Level-1 (L1) trigger and software-based High-Level Trigger (HLT). The former can reduce the events rate from 40 MHz to 100 kHz while the latter can further reduce the events from 100 kHz to 1 kHz.

In terms of the L1 trigger, it includes the L1 calorimeter trigger and L1 muon trigger. The information from calorimeters is fed to the L1 calorimeter trigger composed of the Cluster Processor (CP) for the selection of *tau*-lepton, electron and photon candidates, and the Jet/Energy-sum Processor (JEP) for the selection of jet candidates and calculation of total transverse energy. The information from the triggering chambers of muon spectrometer goes to the L1 muon trigger for the selection of muon candidates. Then the outputs of the L1 calorimeter trigger and L1 muon trigger are further handled by the Central Trigger Processor (CTP). By combining the CTP, L1 topological processor (L1Topo) [55] and some other systems, the final decision of the L1 trigger is made. It's worth mentioning that, the L1Topo can be taken as a collection of dedicated algorithms that process the information of the candidate objects, such as angular distribution and kinematic characteristics, then select the events of interest.

In terms of the HLT trigger, it aims to handle the event stream after the L1 trigger. The dedicated softwares are developed to be more accurate and directional and installed on the computers for making the final trigger decision. The process of the HLT trigger can be regarded as the primary physics analysis but works online. Usually, a higher threshold on the object transverse momentum is set to reduce the event rate. At the end of the HLT trigger, the events are stored on disks at the rate of 1 kHz. Finally, the stored events are analyzed for different physics programs.

2.3 Definition of physics objects

2.3.1 Track and vertex

In the inner detector, the charged particles leave hits in its sensitive cells when passing through. Therefore, the tracks of those charged particles can be restored by fitting the hits with dedicated algorithms. Based on the reconstructed tracks, particle identification and object calibration can be carried out. Vertex is the original point of the collision (primary vertex) or decay (secondary vertex) and quite important in the physics analysis, for example, the *b*-jet tagging, and it can be reconstructed by connecting the tracks backwards to the collision area.

The ATLAS track reconstruction is performed by mainly using the information from SCT and pixel detector. The procedure is presented as follows [56]. The hit clusters are firstly built based on the recorded electrical signals from sensitive cells (pixels and strips). Then the seeds are created from those clusters to serve as the input of a track finder, which

aims to reconstruct tracks in an iterative way. Afterwards, with a combinatorial method [57], many tracks can be built up from each seed. Finally, an algorithm called ambiguity solver is used to evaluate each track and label a score on it.

The track candidates are selected based on the scores that represent the probability of the track being correctly reconstructed. Then they are further filtered by the loose selection criteria defined in [58]. The only exception is the requirement on the minimum p_{T}, which is increased to 500 MeV in Run-2 for adjusting to the higher collision rate. To be used in physics analysis, the events must have at least one primary vertex with at least two associated tracks. It's worth mentioning that, the sum of squared p_{T} of all the associated tracks of the vertex is used as the criteria for primary vertex selection, which requires the largest sum.

2.3.2 Electron and Muon

Electron interact with both the inner detector and ECal. Tracks are measured in the former where a Gaussiansum Fitter [59] is used to fit the tracks in the $\eta - \phi$ space while energy is determined in the latter where the electron candidate can be reconstructed from a seed that is actually an energy deposition cluster matching the fitted tracks in the inner detector. Then a likelihood approach is implemented by combining the information of the clusters and tracks. Finally, electron can be identified by applying selection on the likelihood result.

With the good penetrating ability through matters, muon can go through all the way from the innermost region to the outermost region of the ATLAS detector, and its trajectory is left in both the inner detector and muon spectrometer while a small portion of its energy is deposited in the calorimeter. There are in total four types of muon objects [60] defined and used in physics analysis, which are respectively segmenttagged, combined, calorimetertagged and extrapolated muons. The definition of the muon objects is based on how the muon-related information from different sub-detectors is combined.

Isolation is one of the most conspicuous feature of the electrons and muons that are originated from the events of interest while those from background processes are not isolated with other particles in the $\eta - \phi$ space. Therefore, it will be efficient for background rejection if the isolation of the electrons and muons can be used as an event selection criteria, which is usually defined based on the information from either inner detector track or calorimeter. More specifically, the variable used as the figure of merit of being isolation is defined as the sum of energy or momentum of the objects within the electron- or muon-centred cone.

The working points are defined as different threshold applied on the identification or isolation of electron and muons, thus different working points can result in different signal efficiencies. They are usually selected according to actual requirements of physics analysis. For the electron, the requirements on its transverse momentum and pseudorapidity are larger than 7 GeV and less than 2.5 (excluding 1.37-1.52), while for the muon, they are respectively

larger than 7 GeV and less than 2.7.

2.3.3 Jet

Jet is actually the trace of hadronization process in the calorimeter, taking the shape of a cone with many kinds of hadrons contained, including mainly pions, and also neutrons, protons, kaons and other hadrons. Compared with the electron and muon, the measurement on hadron is lack of accuracy in the ATLAS detector. The dedicated algorithm called anti-k_t [61] is used to reconstruct jet and its key parameter is the radius of the jet cone. There are two methods that are commonly-used in jet reconstruction and calibration, one is called topo-clusters method [62] while the other is the PFlow method [63].

In terms of the topo-clusters method, it aims to find the prominent signal cells against the noises induced by pile-up in the calorimeter, then generate the hadronic topo-clusters, which are calibrated at the EM scale. This kind of calibration can make it have the same response with the electromagnetic showers. Then a weighting scheme called local cluster (LC) is implemented to further correct the hadronic clusters to the right scale.

In terms of the PFlow method, it takes the information from both the inner detector and calorimeter into account. More specifically, it finds the charged particles of the specific topocluster that can match the tracks in the inner detector, then deducts their energy from the topocluster energy. After that, the two parts, tracks of the matched charged hadrons and the residual topocluster after subtraction, are considered as the particle flow objects (PFO). Compared with the topo-clusters method, the PFlow method has better energy resolution.

In the physics analysis of Higgs boson pair search, there are three types of jets used, which are small-radius(R) jets, large-radius(R) and variable-radius(R) track jets. In the analysis where the jets originated from $H \longrightarrow bb$ decay can be separately reconstructed, small-R jets are defined as the PFlow jets with the radius of 0.4 [63]. To reject the jets originated from the backgrounds caused by non-collision process and the noises raised in the calorimeters, the jet cleaning algorithms are used [64] while the jetvertextagger (JVT) is used to reduce the pile-up jets.

The b-jets are the jets initiated from b-quarks and they are the decay products of our signal. Due to the long lifetime of b-hadron, there is an obvious secondary vertex that can be used to tag the b-jets. They should be distinguished from the jets that are initiated from other sources. To do that, the b-tagging algorithm is designed based on a deep neural network (DNN) called DL1r [65, 66], which takes low-level information as its inputs, such as the information extracted from the reconstructed jet, primary vertex, secondary vertex, associated tracks and etc. To further correct the b-tagged jets momentum and improve the invariant mass resolution of two b-jets, the methods μ-in-jet and PtReco [67] are applied on the b-tagged jets.

In the analysis where the Higgs boson is boosted enough so that the jets originated from

$H \longrightarrow bb$ decay can not be separately reconstructed, instead a fat jet should be considered and it can be reconstructed into a large-R jet [68] from the LCTopo jet with the radius parameter set as 1.0. There exist underlying and pile-up events whose products may contaminate the LCTopo jet, so a trimming procedure [69] is then applied on those large-R jet candidates. The idea is to apply the k_t algorithm [70, 71] on the large-R jet again to re-cluster it into several sub-jets with the radius parameter of R_{sub}. Then for each sub-jet, a fraction f_{cut} is defined as the ratio of the sub-jet p_T to the large-R jet p_T. If the f_{cut} of a sub-jet is less than a specific value, the sub-jet is removed. The selection of the parameters R_{sub} and f_{cut} is studied in the Ref. [72] where it's pointed out that $R_{sub} = 0.2$ and $f_{cut} = 5\%$ has the best performance. A combined mass approach [73] is used to perform the large-R jet mass calculation. Before talking about b-tagging of large-R jets, the variable-R track jet must be introduced. Again the anti-k_t algorithm is used to reconstruct this kind of jet in the inner detector, but instead of a fixed radius parameter, the p_T-dependent radius parameter defined as $R(p_{\mathrm{T}}) = \rho/p_{\mathrm{T}}$ is set for the algorithm. Here ρ is a constant (30 GeV), which leads to the varying radius parameter from 0.02 to 0.4 [74]. After the reconstruction of the variable-R track jets, a method called ghost association [75, 76] is used to allocate the track jets to large-R jets. Then b-tagging should be firstly applied on those associated variable-R track jets of the large-R jet, and order the b-tagged ones by p_T. Finally the large-R jet b-tagging is performed by requiring at least two leading b-tagged variable-R track jets based on the $MV2c10$ method [77].

In the search for Higgs boson pair production, the three types of jets are selected with different requirements. For the small-R jet, the cuts $p_{\mathrm{T}} > 20$ GeV and $\eta < 2.5$ are applied and usually the DL1r 77% working point is used. For the large-R jet, the requirement on p_{T} is quite large $p_{\mathrm{T}} > 300$ GeV while the requirement on η is still $\eta < 2.5$. For the variable-R track jet, the p_{T} is required to be low $p_{\mathrm{T}} > 10$ GeV and still $\eta < 2.5$ is applied, the $MV2c10$ 70% working point is usually used.

2.3.4 Overlap removal

There exist the chances that the reconstructed objects may geometrically overlap with each other, which causes ambiguity in the physics analysis. To avoid that, an overlap removal procedure is applied on the candidate objects. The following steps are from the standard overlap removal tool with recommended working point.

- Reject the electron with smaller p_{T} if two electrons share the same track.
- Reject τ_{had} against electron if $\Delta R < 0.2$.
- Reject τ_{had} against muon if $\Delta R < 0.2$.
- Reject muon against electron if the muon is calorimeter-tagged and shares the inner detector track with the electron.

- Reject electron against muon if they share inner detector track.
- Reject jet against electron if $\Delta R < 0.2$.
- Reject electron against jet if $\Delta R < 0.4$.
- Reject jet against muon if $N_{\text{track}} < 3$ ($p_{\text{T,track}} > 500\,\text{MeV}$), and the muon inner detector track is ghost-associated to the jet or $\Delta R < 0.2$.
- Reject muon against jet if $\Delta R < 0.4$.

2.3.5 Missing transverse energy

The missing transverse momentum $\vec{p}_{\text{T}}^{\,\text{miss}}$ is defined as the negative vectorial sum of the transverse momenta of other reconstructed objects [78], such as electrons, muons, jets, photons, $\tau_{had-vis}$. In addition, soft terms from tracks that are matched to the primary vertex but not associated with the higher-p_{T} reconstructed objects, are handled by the track-based soft term (TST) algorithm.

The $\vec{p}_{\text{T}}^{\,\text{miss}}$ is calculated as the following equation:

$$\vec{p}_{\text{T}}^{\,\text{miss}} = -\sum \vec{p}_{\text{T}}^{\,e} - \sum \vec{p}_{\text{T}}^{\,\mu} - \sum \vec{p}_{\text{T}}^{\,\gamma} - \sum \vec{p}_{\text{T}}^{\,\tau_{had-vis}} - \sum \vec{p}_{\text{T}}^{\,\text{jet}} - \sum \vec{p}_{\text{T}}^{\,\text{soft}}, \tag{2.1}$$

The missing transverse energy ($E_{\text{T}}^{\text{miss}}$) is defined as the magnitude of missing transverse momentum vector. In the search for double-Higgs decaying to $b\bar{b}\tau^+\tau^-$, the neutrinos from the τ-leptons decays contribute to the majority of $E_{\text{T}}^{\text{miss}}$.

Chapter 3

Search for Higgs Boson Pair Production

This chapter summarizes the legacy search for the non-resonant Higgs boson pair production in the $HH \longrightarrow b\bar{b}\tau^+\tau^-$ channel. The search is performed on a proton-proton collision dataset with an integrated luminosity of 140 fb^{-1} at $\sqrt{s}$ = 13 TeV collected by the ATLAS detector at the LHC. It is actually an updated search based on previous ATLAS result in the same channel [79]. The general identification of objects, strategy for triggering, event selection and background estimation in the signal-enriched regions remains the same. However, an optimized classification of the selected events has been introduced to improve the sensitivity to the modifier κ_λ of the Higgs boson self-coupling and also the modifier κ_{2V} of the quartic $HHVV$ ($V = W, Z$) coupling.

The classification is implemented as follows. Just like previous analysis, this search is also divided into two sub-channels: the $b\bar{b}\ \tau_{\text{had}}\tau_{\text{had}}$ and the $b\bar{b}\ \tau_{\text{lep}}\tau_{\text{had}}$, with the latter further split into two categories based on the trigger choice: single-lepton triggers (SLT) and lepton-plus-$\tau_{\text{had-vis}}$ triggers (LTT). For this round of the analysis, events in each sub-channel are then categorized into dedicated ggF and VBF signal regions (SRs) using multivariate techniques to separate the signal contributions from the ggF and VBF production modes. Additionally, the ggF regions are split into low- and high-mass SRs depending on the invariant mass of the HH system in each event. For the fit discriminant, optimized multivariate scores are used in each SRs.

In addition to the updated classification strategy, there are also other improvements to this analysis compared to previous publication. Monte Carlo predictions are updated to better describe main backgrounds of top-quark pair production ($t\bar{t}$) and Z boson production in association with heavy flavor quarks, which lead to a more accurate modeling of these processes and enhance the statistical power of the simulation. The event selection for the measurement of the background from Z boson production in association with heavy flavor quarks is adapted to improve the consistency of the kinematic properties of this process with the signal-enriched regions. The binning strategy of the fit discriminant is updated. Finally,

the 95% confidence-level upper limit on the Higgs-boson pair production cross-section, and the 95% confidence interval of both κ_λ and κ_{2V} are presented as the final results.

3.1 Data and Monte Carlo samples

3.1.1 Data samples

The results presented here are based on the full Run-2 proton-proton collision data at a centre-of-mass energy of $\sqrt{s}$ =13 TeV, collected by the ATLAS detector at the LHC between 2015 and 2018, corresponding to an integrated luminosity of $\mathscr{L}$ = 140.1 ± 1.2 fb^{-1} [80], where the selected data events are required to have all relevant components of the ATLAS detector in good working condition according to the ATLAS Good-Run-List (GRL).

3.1.2 Monte Carlo samples

Monte Carlo (MC) simulated events are used to model SM background production, SM *HH* signal production, and BSM *HH* signal production. The events were passed through the full ATLAS detector simulation [81] based on GEANT4 [82]. The effects of pile-up in the same and neighbouring bunch crossings were modelled by overlaying each hard-scatter event with minimum-bias events, simulated using the soft quantum chromodynamics (QCD) processes of PYTHIA 8.186 [83] with a set of tuned parameters called the A3 tune [84] and the NNPDF2.3LO [85] parton distribution functions (PDFs). The EVTGEN program [86] was used to model the decays of bottom and charm hadrons in all samples of simulated events, except those generated using SHERPA [87]. The samples generated with SHERPA used the bottom- and charm-hadron decay model implemented within the generator. The simulated events were processed through the same reconstruction algorithms as the data. For all samples containing a SM Higgs boson, its mass was fixed to 125 GeV. The same mass value is used in the calculation of the Higgs boson decay branching fractions and in the calculation of the single-Higgs-boson and SM *HH* production cross-sections. Unless otherwise specified, the order of the cross-section calculation refers to the expansion in the strong coupling constant (α_s).

3.1.2.1 Signal samples

There are two types of *HH* signal samples simulated, which are respectively from ggF and VBF processes.

Nominal ggF signal samples (for $\kappa_\lambda = 1.0, 10.0$) were simulated with the POWHEG BOX v2 generator [23] at next-to-leading order (NLO) with finite top-quark mass, and using the PDF4LHC15_NLO_30_PDFAS (code 90400 in the LHAPDF database [88]) PDF set [89].

Parton showers and hadronization were simulated using PYTHIA 8.244 [83] with the A14 tune [90, 91] and the NNPDF2.3LO PDF set. ggF HH events with different κ_λ values (in the interval $\kappa_\lambda \in [-20, 20]$) can be provided by reweighting either of the two available samples using a reweighting approach, which provides per-event weights, depending on the targeted κ_λ value and on the truth m_{HH}. The normalization of the SM process is set to the SM ggF di-Higgs cross-section, $\sigma_{\text{ggF}} = 31.05\,\text{fb}$, calculated at NNLO FTApprox [92], times the $b\bar{b}\tau^+\tau^-$ branching ratio, $\sigma_{\text{ggF}} \times \mathcal{B}(b\bar{b}\tau^+\tau^-) = 31.05\,\text{fb} \times 0.0730562561 = 2.268\,396\,7\,\text{fb}$. The normalization of the BSM process is derived from a 2$^{\text{nd}}$ order Polynomial function of κ_λ [93]. Alternative ggF signal samples (for $\kappa_\lambda = 1.0$, 10.0) to assess the parton showering uncertainty were simulated using POWHEG BOX v2 generator [23] at NLO, interfaced to HERWIG 7 [94, 95].

The MADGRAPH5_aMC@NLO 2.7.3 [26] generator with the NNPDF3.0NLO [96] PDF set is used to generate the nominal VBF signal samples at LO. Parton showering and hadronization were performed using PYTHIA 8.244 with the A14 tune and the NNPDF2.3LO PDF set. VBF samples have been produced for different values of the coupling modifiers κ_λ, κ_{2V}, and κ_V, where the SM scenario corresponds to $\kappa_\lambda = \kappa_V = \kappa_{2V} = 1$. A linear combination of six (κ_λ, κ_{2V}, κ_V) basis samples is used to derive distributions for a finer granularity of κ_{2V} values. These six basis samples were chosen to lie in the region where the analysis is expected to be sensitive, in order to avoid large statistical uncertainties in the interpolation resulting from the reweighting. Given the large pseudo-rapidity gap between the two VBF jets is a distinct feature for VBF jets, this variable provides a good handle for the separation of ggF and VBF HH processes. The normalization of the SM process is set to the SM VBF HH cross-section, $\sigma_{\text{VBF}} = 1.726\,\text{fb}$ calculated at next-to-next-to-next-to-leading order (N^{3}LO) in QCD in the limit of no partonic exchange between the two protons [97], times the $b\bar{b}\tau^+\tau^-$ branching ratio, $\sigma_{\text{VBF}} \times \mathcal{B}(b\bar{b}\tau^+\tau^-) = 1.726\,\text{fb} \times 0.0730562561 = 0.126\,095\,098\,\text{fb}$. The normalization of the BSM processes is derived from a 2$^{\text{nd}}$ order polynomial function of κ_{2V} [93]. This was done by using the generator cross-sections at LO with κ_{2V} variations scaled to N3LO with the SM k-factor, followed by a 2$^{\text{nd}}$ order polynomial fit to these values. For the parton shower variations, alternative VBF signal samples were generated at three sets of coupling values (κ_λ, κ_{2V}, κ_V) = (1, 1, 1), (1, 0, 1), (10, 1, 1). The alternative samples were generated at LO using the MADGRAPH5_aMC@NLO 2.7.3 [26] generator with the NNPDF3.0NLO [96] PDF set and interfaced to HERWIG 7 [94, 95].

3.1.2.2 Background samples

The $t\bar{t}$ production and single top-quarks production in the Wt , s and t-channels are simulated using the POWHEG BOX v2 generator [98–100]. The NNPDF30NLO [101] PDF set is used. The events are interfaced to PYTHIA 8 version 8.230 [102] for the parton shower and hadronization with the A14 set of tuned parameters [91, 103] and the NNPDF23LO [104]

PDF set. The EvtGen v1.6.0 program [105] is used to model the properties of the bottom and charm hadron decays. For all top processes, top-quark spin correlations are preserved (for t-channel production, top quarks are decayed using MadSpin [106]). The top-quark mass is set to 172.5 GeV. The NLO $t\bar{t}$ production cross section is corrected to the theory prediction calculated at NNLO+NNLL. For single top-quark processes, the cross sections were corrected to the theory predictions calculated at NLO. The $t\bar{t}$-Wt interference is handled using the diagram removal scheme. Moreover, an additional $t\bar{t}$ dilepton sample is used in this round of the analysis to increase the overall $t\bar{t}$ statistics. More specifically, a dedicated dilepton $t\bar{t}$ sample (DSID 410472) produced with Powheg + Pythia 8 was combined with the non-all-hadronic $t\bar{t}$ sample (DSID 410470), provided the dilepton events are removed from the latter.

Events containing W or Z bosons produced in association with jets are simulated using the Sherpa 2.2.11 [87] generator using NLO matrix elements for up to two partons, and LO matrix elements for up to five partons calculated with the Comix [107] and OpenLoops [108–110] libraries. They are matched with Sherpa parton shower using the MEPS@NLO prescription [111–114] in conjunction with a set of tuned parameters developed by the Sherpa authors. The NNPDF3.0nnlo set of PDFs [96] is used. The choice of Sherpa 2.2.11 configuration (instead of Sherpa 2.2.1 which was adopted in the previous analysis round and known to underestimate the heavy-flavor hadron production fractions) is based on Ref. [115]. Sherpa 2.2.11 shows a significant MC statistical uncertainty improvement for b-, c- and light-flavor and overall, a gain with a factor of ~ 2 in the number of events.

Diboson processes with one of the bosons decaying hadronically and the other leptonically are simulated using the Sherpa version 2.2.1 [87] generator. The NNPDF3.0nnlo set of PDFs [96] is used in conjunction with dedicated parton shower tuning developed by the Sherpa authors. The generator NLO cross-sections are used. Production of W and Z bosons in association with a top-quark pair, ttV, is simulated using Sherpa version 2.2.1 with multi-leg NLO merging for the ttZ production and using Sherpa version 2.2.8 at NLO for the ttW production. The NNPDF3.0nnlo set of PDFs [96] is used in conjunction with dedicated parton shower tuning developed by the Sherpa authors. The most accurate NLO generator cross-sections are used.

SM single Higgs boson production is included in the analysis as part of the background processes, and is described as follows.

SM Higgs production in association with a top-quark pair, ttH, is simulated using the Powheg Box generator [98–100]. The NNPDF30NLO PDF set is used. The events are interfaced to Pythia 8 version 8.230 [102] for the parton shower and hadronization with the A14 set of tuned parameters [91, 103] and the NNPDF23LO PDF set [104]. The EvtGen program [105] is also used. The cross-section is set to ttH production NLO calculations [116]. The Higgs boson production in association with a Z boson, ZH, with the Higgs boson

decaying to bb or $\tau\tau$, is included in the analysis using four samples. The $qqZH(Z \to ll, H \to bb)$, $ggZH(Z \to ll, H \to bb)$ (where "l" includes all leptons e, μ, τ) and $qqZH(Z \to all, H \to \tau\tau)$, $ggZH(Z \to all, H \to \tau\tau)$ are simulated using POWHEG BOX v2. The NNPDF3.0NNLO set of PDFs [96] is used. The events are interfaced with PYTHIA 8 version 8.212 using the AZNLO tune [117] and the CTEQ6L1 PDF set [118]. The EvtGen program [105] is also used. The cross section is set to the NNLO(QCD)+NLO(EW) calculations for $qqZH$ and to the NLO+NLL in QCD for $ggZH$. The Higgs boson production in association with a W boson, WH, with the Higgs boson decaying to bb or $\tau\tau$, is included in the analysis using four samples (separate samples for W^+ and W^-). The $W^{\pm}H(W \to lv, H \to bb)$, $W^{\pm}H(W \to all, H \to \tau\tau)$ are simulated using POWHEG BOX v2. The NNPDF3.0NNLO set of PDFs [96] is used. The events are interfaced with PYTHIA 8 version 8.212 using the AZNLO tune [117] and the CTEQ6L1 PDF set [118]. The EvtGen program [105] is also used. The cross-section is set to the NNLO(QCD)+NLO(EW) calculations.

The ggF Higgs boson production with the Higgs boson decaying to $\tau\tau$ is simulated using POWHEG BOX v2. The NNPDF3.0NNLO set of PDFs [96] is used. The events are interfaced with PYTHIA 8 version 8.212 using the AZNLO tune [117] and the CTEQ6L1 PDF set [118]. The EvtGen program [105] is also used. The cross section is set to the N3LO(QCD)+NLO(EW) calculations [116]. The VBF Higgs boson production with the Higgs boson decaying to $\tau\tau$ is simulated using POWHEG BOX v2. The NNPDF3.0NNLO set of PDFs [96]. The events are interfaced with PYTHIA 8 version 8.212 using the AZNLO tune [117] and the CTEQ6L1 PDF set [118]. The EvtGen program [105] is also used. The cross section is set to the NNLO(QCD)+NLO(EW) calculations [116].

All MC samples are passed through the full GEANT4 [119, 120] simulation of the ATLAS detector and are reconstructed with the same software as used for data. Additional samples produced with alternative generators and settings are used to estimate systematic uncertainties in the event modelling.

3.2 Event selections

3.2.1 Overview

The analysis is divided into two sub-channels depending on the di-τ decay mode as depicted in Figure 3.1. The $b\bar{b}$ $\tau_{\text{had}}\tau_{\text{had}}$ sub-channel targets events with two oppositely charged $\tau_{\text{had-vis}}$ and two b-jets, whereas the $b\bar{b}$ $\tau_{\text{lep}}\tau_{\text{had}}$ sub-channel, which gets further split into two categories based on the trigger choice, targets events with an electron or muon, an oppositely charged $\tau_{\text{had-vis}}$ and two b-jets. In both sub-channels, the two b-jets are requested to pass the 77% efficiency working point. The description of the trigger configuration for the $b\bar{b}$ $\tau_{\text{lep}}\tau_{\text{had}}$ sub-channel, with single-lepton triggers (SLT) and lepton-plus-$\tau_{\text{had-vis}}$ triggers

(LTT), and for the $b\bar{b}\,\tau_{\text{had}}\tau_{\text{had}}$ sub-channel, with combination of single-$\tau_{\text{had-vis}}$ triggers (STT) and di-$\tau_{\text{had-vis}}$ triggers (DTT) are presented in Section 3.2.2.

Further, events in each sub-channel are categorized into ggF and VBF signal regions (SRs) by using dedicated BDTs, and the ggF regions are in turn split into low- and high-mass SRs depending on the invariant mass of the HH system in each event, by using a mass cut of 350 GeV, as shown in Figure 3.1. A control region (Z+HF CR) is also defined in order to determine the normalization of the background in which a Z boson is produced in association with one or more jets initiated by heavy-flavor quarks (Z+HF background), by using single-lepton and di-lepton triggers and specific event selection to finally extract the $m_{\ell\ell}$ shape as parameter of interest. This trigger strategy has been adopted from the $bb\ell\ell$ analysis [121].

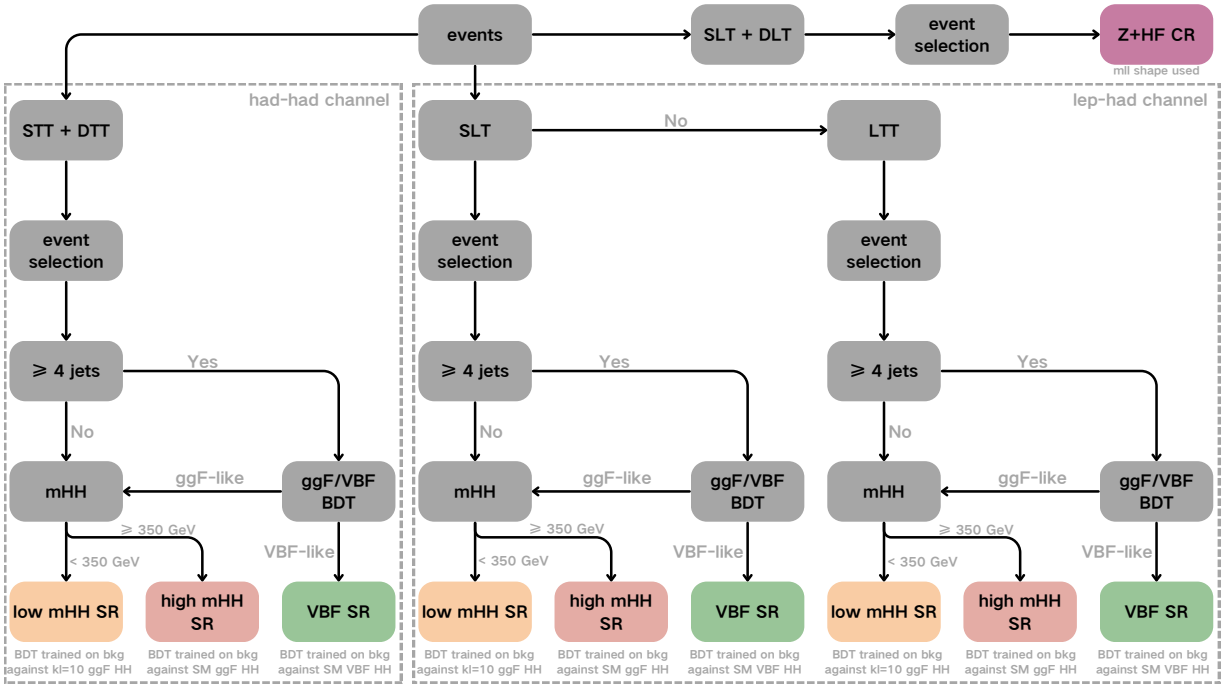


Figure 3.1: Sketch depicting the analysis strategy starting with the triggers and event selection for the $b\bar{b}\,\tau_{\text{had}}\tau_{\text{had}}$ and $b\bar{b}\,\tau_{\text{lep}}\tau_{\text{had}}$ sub-channels, followed by a BDT approach defining the orthogonality between ggF and VBF signal regions. The ggF signal region is subsequently split based on a m_{HH} cut of 350 GeV to improve the sensitivity of enhanced κ_λ. Furthermore, a control region (Z+ hf) is defined based on various single- and di-lepton triggers and event selection with the $m_{\ell\ell}$ shape used as parameter of interest. STT: single $\tau_{\text{had-vis}}$ triggers; DTT: di-$\tau_{\text{had-vis}}$ triggers; SLT: single lepton triggers; LTT: lepton+$\tau_{\text{had-vis}}$ triggers; DLT: di-lepton triggers.

In the following, we give a brief description of the triggers used in each sub-channel and focus more on the updated event categorization (Section 3.2.4) besides the redefinition of the (Z+HF CR) phase space (Section 3.2.5) with respect to the previous round of the analysis [79].

3.2.2 Trigger selection

The description of the trigger selection hasn't changed since the previous round of the analysis.

For the $\tau_{\text{lep}}\tau_{\text{had}}$ sub-channel, events are recorded using a combination of single-lepton triggers (SLTs) and lepton-plus-$\tau_{\text{had-vis}}$ triggers (LTTs). The SLTs require an electron or muon to be reconstructed at the HLT with an E_{T} threshold that ranges from 24 GeV to 26 GeV for electrons and a p_{T} threshold that ranges from 20 GeV to 26 GeV for muons, depending on the data-taking period. The LTTs require that an electron with $p_{\text{T}} > 17$ GeV or a muon with $p_{\text{T}} > 14$ GeV in addition to a $\tau_{\text{had-vis}}$ with $p_{\text{T}} > 25$ GeV is reconstructed at the HLT. The LTTs used to collect lepton-plus-$\tau_{\text{had-vis}}$ events with $\tau_{\text{had-vis}}$ $p_{\text{T}} < 35$ GeV had additional requirements at the L1 trigger, requiring the presence of either an additional jet with $E_{\text{T}} > 25$ GeV or two additional jets with $E_{\text{T}} > 12$ GeV. The electron-plus-$\tau_{\text{had-vis}}$ triggers that require an additional jet with $E_{\text{T}} > 25$ GeV was only used to select an event if the electron-plus-$\tau_{\text{had-vis}}$ triggers that require two additional jets with $E_{\text{T}} > 12$ GeV did not select the event. For any trigger to select an event, based on the presence of a muon, the muon must have $|\eta| < 2.5$. In order to select events near the trigger efficiency plateaus where the trigger efficiencies are well modelled, the offline electrons, muons and $\tau_{\text{had-vis}}$ objects are required to be within $\Delta R = 0.07$, $\Delta R = 0.1$ and $\Delta R = 0.2$ of the corresponding objects at the HLT, respectively. Minimum p_{T} requirements are applied to the offline objects, and these are 1 GeV above the thresholds for electrons and muons at the HLT, 5 GeV above the thresholds for $\tau_{\text{had-vis}}$ at the HLT, and 80 GeV(45 GeV) for jets with an L1-trigger E_{T} threshold of 25 GeV(12 GeV). Events which pass the offline SLT lepton p_{T} requirements are not considered for the LTT. This ensures that there is no overlap between the SLT and LTT categories. These two categories are analyzed separately.

For the $\tau_{\text{had}}\tau_{\text{had}}$ sub-channel, events are recorded using a combination of single-$\tau_{\text{had-vis}}$ triggers (STTs) and di-$\tau_{\text{had-vis}}$ triggers (DTTs). The STTs accept events with at least a single $\tau_{\text{had-vis}}$ at the HLT with a minimum p_{T} between 80 GeV and 160 GeV, depending on the data-taking period. The DTTs select events with at least a pair of $\tau_{\text{had-vis}}$ reconstructed at the HLT, with minimum p_{T} of 35 GeV (25 GeV) for the (sub-)leading $\tau_{\text{had-vis}}$, where the (sub-)leading $\tau_{\text{had-vis}}$ is defined as the $\tau_{\text{had-vis}}$ with the (second-)highest p_{T} in the event. From the 2016 data-taking period onward, additional requirements were applied in the L1 trigger to reduce the DTT rates. During 2016 data-taking, an additional jet with $E_{\text{T}} > 25$ GeV was required. For 2017 (2018) data-taking, if two offline jets with $p_{\text{T}} > 45$ GeV are found, then a trigger is used that requires two additional jets with $E_{\text{T}} > 12$ GeV (and $|\eta| < 2.3$) at L1, otherwise another trigger is used that requires one additional jet with $E_{\text{T}} > 25$ GeV and the $\tau_{\text{had-vis}}$ to be reconstructed within $\Delta R = 2.8$ of each other[1]. In order to select events near

[1]This dependence of the online selection on the offline requirements ensures that both triggers will not be used for the same events, which avoids the need to study their combined efficiency.

the trigger efficiency plateaus where the efficiencies are well modelled, the offline $\tau_{\text{had-vis}}$ are required to be within $\Delta R = 0.2$ of the corresponding HLT $\tau_{\text{had-vis}}$ objects and a minimum offline p_{T} requirement is applied to $\tau_{\text{had-vis}}$ and to the jets. The offline p_{T} thresholds for the $\tau_{\text{had-vis}}$ range between 100 GeVand 180 GeVfor the STTs, and are 40 GeV(30 GeV) for the (sub-)leading $\tau_{\text{had-vis}}$ for the DTTs. Additional offline requirements for the DTTs are either that two additional jets with $p_{\text{T}} > 45$ GeV are present in the event, or that a jet with $p_{\text{T}} > 80$ GeV is present in the event and the $\tau_{\text{had-vis}}$ are reconstructed within $\Delta R = 2.5$ of each other. For events that pass both the STTs and DTTs, the offline requirements used for the STTs are applied. Events passing the $\tau_{\text{had}}\tau_{\text{had}}$ event selection are analyzed together.

3.2.3 Event selection

In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, after the trigger-based selection described in Section 3.2.2, exactly one 'loose' electron or 'loose' muon, an oppositely charged $\tau_{\text{had-vis}}$, and exactly two b-tagged jets are required. The selected electron (muon) must also pass a tight (medium) identification requirement with an efficiency of around 80% (97%). Events are required to have $m_{\tau\tau}^{\text{MMC}} > 60$ GeV [122], where $m_{\tau\tau}^{\text{MMC}}$ is calculated using the four-momenta of the electron or muon, the $\tau_{\text{had-vis}}$ and the $\mathbf{p}_{\text{T}}^{\text{miss}}$. The b-tagged jet pair invariant mass (m_{bb}) is required to be less than 150 GeV to reject $t\bar{t}$ background events and to allow for the definition of a $t\bar{t}$-enriched region which is used in the estimation of $t\bar{t}$ backgrounds. A $\tau_{\text{had-vis}}$ with $p_{\text{T}} > 20$ GeV and $|\eta| < 2.3$ is required in the SLT category, and a $\tau_{\text{had-vis}}$ with $p_{\text{T}} > 30$ GeV, or higher if required by the trigger, and $|\eta| < 2.3$ is required in the LTT category. In both categories, the (sub-)leading b-tagged jet must have $p_{\text{T}} > 45$ (20) GeV, in addition to any trigger-dependent requirements. The event selection leading to the $\tau_{\text{lep}}\tau_{\text{had}}$ signal region categories is summarized in Table 3.1.

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, on top of the trigger selection described in Section 3.2.2, further selection requirements are applied to form the signal region. Events are required to contain exactly two τ-leptons passing the 'loose' identification criteria and having opposite-sign charges and at least two jets, exactly two of which should be b-tagged (passing the DL1r 77% working point). Events with any additional leptons (electrons or muons) are vetoed, and the (sub)-leading b-tagged jet is required to have $p_{\text{T}} > 45\,(20)$ GeV. The invariant mass of the τ-lepton pair, $m_{\tau\tau}^{\text{MMC}}$, is estimated from the four-momenta of the $\tau_{\text{had-vis}}$ and the $\mathbf{p}_{\text{T}}^{\text{miss}}$ using the Missing Mass Calculator (MMC) [122], which assumes that the $\mathbf{p}_{\text{T}}^{\text{miss}}$ is exclusively from the neutrinos produced in the τ-lepton decays. To reject background from low-mass Drell-Yan events, $m_{\tau\tau}^{\text{MMC}}$ is required to be above 60 GeV. The event selection leading to the $\tau_{\text{had}}\tau_{\text{had}}$ signal region category is also summarized in Table 3.1.

Table 3.1: Summary of the event selections, shown separately for events that are selected by different triggers. In cases where pairs of reconstructed objects of the same type are required, thresholds for the (sub-)leading p_{T} object are given outside (within) parentheses where different event selection thresholds are applied. When the selection depends on the year of data-taking, the possible values of the requirements are separated by commas, except for the jet selection in the LTT and DTT triggers, which use multiple selection criteria. The trigger p_{T} thresholds shown are applied to the offline physics objects that are matched to the corresponding trigger objects.

$\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ category		$\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ categories	
STT	DTT	SLT	LTT
e/μ selection			
No loose e/μ	No loose e/μ	Exactly one loose e/μ	
		e (μ) must be tight (medium and have $\lvert\eta\rvert < 2.5$)	
		$p_{\mathrm{T}}^{e} > 25, 27$ GeV	18 GeV $< p_{\mathrm{T}}^{e} <$ SLT cut
		$p_{\mathrm{T}}^{\mu} > 21, 27$ GeV	15 GeV $< p_{\mathrm{T}}^{\mu} <$ SLT cut
$\tau_{\text{had-vis}}$ selection			
Two loose $\tau_{\text{had-vis}}$	Two loose $\tau_{\text{had-vis}}$	One loose $\tau_{\text{had-vis}}$	
		$\lvert\eta\rvert < 2.3$	
$p_{\mathrm{T}} >$ 100, 140, 180 (25) GeV	$p_{\mathrm{T}} > 40$ (30) GeV		$p_{\mathrm{T}} > 30$ GeV
Jet selection			
		≥ 2 jets with $\lvert\eta\rvert < 2.5$	
Leading jet $p_{\mathrm{T}} > 45$ GeV	Trigger dependent	Leading jet $p_{\mathrm{T}} > 45$ GeV	Trigger dependent
Event-level selection			
Trigger requirements passed			
Collision vertex reconstructed			
$m_{\tau\tau}^{\mathrm{MMC}} > 60$ GeV			
Opposite-sign electric charges of $e/\mu/\tau_{\text{had-vis}}$ and $\tau_{\text{had-vis}}$			
Exactly two b-tagged jets			
		$m_{bb} < 150$ GeV	

3.2.4 Event categorization

In contrast to the previous full Run-2 analysis, here the events passing all the selection criteria described above are further divided into three mutually exclusive categories, as shown in the sketch of Figure 3.1, resulting in three signal regions per sub-channel. The events forming these signal regions are then used for the fit of the MVA discriminant distributions.

If an event has at least four jets, it goes through a ggF/VBF BDT classifier which aims to optimally separate events originating from ggF and VBF production modes. Then, if the event passes a certain ggF/VBF BDT working point, it is considered VBF-like and falls into the dedicated VBF signal region. On the other hand, if an event has less than four jets or fails the ggF/VBF BDT cut, it is considered ggF-like and is further categorized based on the invariant mass of the di-Higgs system, m_{HH}. A low-mass ggF signal region is defined by $m_{HH} < 350$ GeVand targets BSM signals with $\kappa_\lambda \neq 1$ values that are enhanced in that region. Finally, events with $m_{HH} > 350$ GeVare categorized in a high-mass ggF region targeting SM-like scenarios. The ggF/VBF BDT is trained to discriminate between the SM ggF HH and the SM VBF HH signal processes, with the former as signal and the latter as background. For the ggF/VBF BDT working point, the cut value that gives the best performance is selected. The threshold of 350 GeVis chosen as such to target tight constraints on κ_λ while having large enough sample sizes in the low-mass region and retaining the SM-like HH signal sensitivity. Also, the destructive interference between the two ggF LO Feynman diagrams becomes maximal at that value.

Another BDT classifier is trained in each signal region using different HH signal hypotheses each time to separate HH signal events from background processes. It is used as the final discriminant of this analysis. In the VBF region, a BDT is trained using the SM VBF HH signal. While SM ggF HH signal is used in the high-mass BDT training, the low-mass region BDT is trained on the ggF κ_λ =10 signal which is enhanced there.

3.2.5 Z+HF control region

The cross-section of Z boson production in association with heavy flavor (b, c) jets is known to be not well predicted by the Sherpa MC and so these processes are normalized to data in a control region. Since the production of jets is independent of the decay mode of the Z boson, $Z \longrightarrow \mu\mu/ee$ + heavy flavor jets are selected as this provides a high purity sample that is orthogonal to the signal regions selection and can be included in the final fit to determine the Z+HF normalization from data. For this round of the analysis, the Z+HF control region phase space was redefined with respect to the previous round.

The $Z + HF$ control region phase space differs significantly from the signal region phase spaces. For example, the SRs select τ-leptons with higher p_{T} and have in addition a leading b-jet p_{T} cut of at least 45 GeV. This results in a generally harder V p_{T} spectrum of the SRs

with respect to the $Z + HF$ control regions. In the previous round of the analysis, this did not matter too much, because the data/MC modelling was flat over V p_{T}. However, the new SHERPA 2.2.11 samples show a significant slope in the data/MC modelling as a function of V p_{T}.

To not rely too much on the phase-space extrapolation uncertainties between SR and CR it was decided to adjust the CR cuts to match closer the SR cuts. For this round of the analysis, the events falling in the redefined control region are selected as follows:

- Events selected with $bb\ell\ell$ trigger selection using single-lepton and di-lepton triggers (see Section 3.1 of Ref. [121]);
- Exactly two muons or two electrons with opposite-sign charges;
- Exactly two b-tagged jets (using DL1r tagger and 77% working point);
- 75 GeV < $m_{\ell\ell}$ < 110 GeV (select Z mass peak);
- m_{bb} < 40 GeV or m_{bb} > 210 GeV (to veto Higgs mass peak and to ensure orthogonality to $bb\ell\ell$ signal region);
- leading b-jet p_{T} > 45 GeV;
- lepton p_{T} > 40 GeV.

When compared to the previous round, the lepton p_{T} cut was raised to 40 GeVand a leading b-jet p_{T} cut of at least 45 GeV was introduced. It has been checked that the reconstructed V p_{T} in the case of the di-τ system after the Missing Mass Calculator aligns with the true di-τ p_{T}, which motivates the choice of the same cuts for electrons and muons, where this is the case. This re-definition brings the V p_{T} shapes of the SR and CR closer, as shown in the direct comparison in Figure 3.2. In addition, this tighter cut also reduces the overlap of the CR with the $bb\ell\ell$ CR and might facilitate combinations given the usage of different samples and different b-tagging strategies.

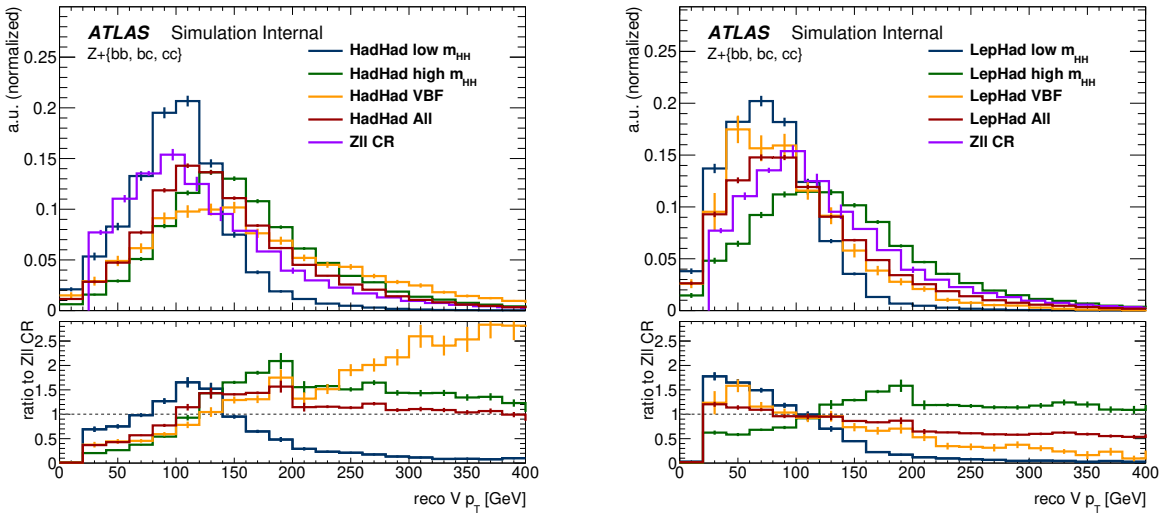


Figure 3.2: Comparison of V p_{T} shapes between the various SRs and the Z CR in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel (left) and the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel (right).

Using the new Z CR and updated Sherpa 2.2.11 prediction, the normalization factor for Z+hf rises to around 1.26, similar to the factor 1.3 from the previous round of the analysis. It has been studied that the tighter cuts in the Z CR and the resulting lower statistics have no impact on the performance of the analysis. The constraining powers on Z+hf and $t\bar{t}$ do get lower when moving to the tighter cuts, as expected but without any impact on the signal strength or κ limits.

3.3 Background estimation and validation

3.3.1 Background estimation strategy

The background estimation follows the same methods used in the previous analysis round for both channels [79].

The simulated event samples summarized in Section 3.1.2 are used to model all background processes, except for processes with fake-τ_{had} which are estimated using data-driven techniques, as discussed below. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, all fake-τ_{had} backgrounds from $t\bar{t}$ and multi-jet processes are estimated using an inclusive fake-factor method. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the multi-jet background is estimated using a data-driven fake-factor method and the $t\bar{t}$ background with fake-τ_{had} is estimated using a scale-factor method with scale-factors derived from data to correct the MC prediction. Separate estimation techniques are used for fake τ_{had} from multi-jet and $t\bar{t}$ processes, since a combined approach is difficult in the $\tau_{\text{had}}\tau_{\text{had}}$-channel where multi-jet features two fake-τ_{had} while typically only a single fake-τ_{had} is present in $t\bar{t}$. Moreover, the presence of HLT τ_{had} identification due to the $\tau_{\text{had}}\tau_{\text{had}}$ trigger selection and the requirement of at least one loose τ_{had} in the derivation skim introduces difficulties in combined estimation procedures. The larger uncertainties, as compared to a combined approach, are not causing any degradation of the results due to the statistical limitation of Run-2 analysis. Also, the fake-τ_{had} background is not a dominant process limiting the sensitivity, given that the contribution is small in most sensitive bins of the analysis, this can be seen as a more conservative approach of the fake estimation.

The $t\bar{t}$ with true-τ_{had} and Z+HF templates are taken from the MC prediction but their normalizations are derived from data as included as freely floating parameters in the final fit. Events with electrons or muons that are misidentified as τ_{had} objects, dominantly coming from the $t\bar{t}$ production, represent a minor background in the analysis and they are estimated from simulation. This background is treated together with the $t\bar{t}$ events containing true-τ_{had} objects.

In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, a combined fake-factor method is used to estimate multi-jet and $t\bar{t}$ backgrounds with fake-$\tau_{\text{had-vis}}$. A schematic depiction of this method is shown in Figure 3.3.

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the multi-jet and $t\bar{t}$ backgrounds with fake-τ_{had} are estimated

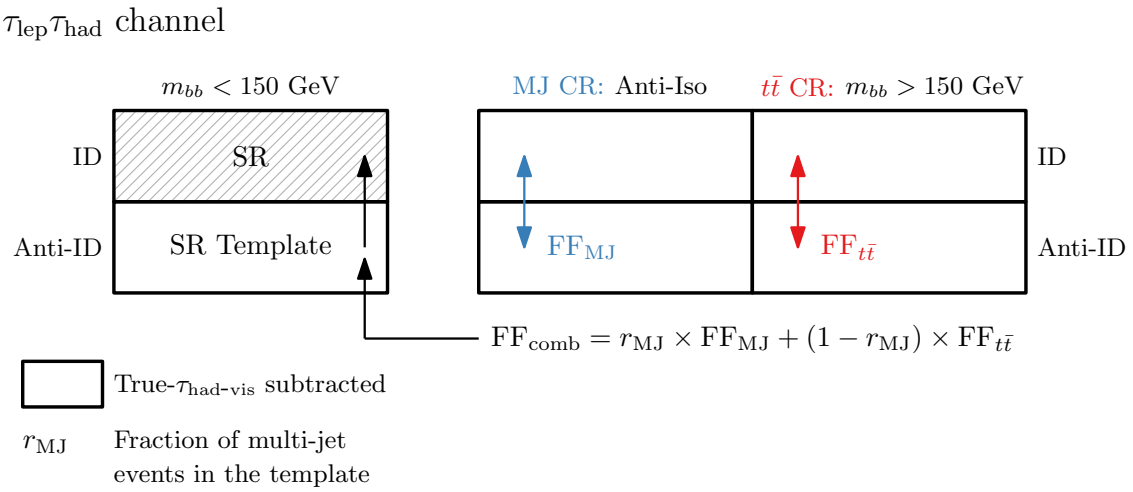


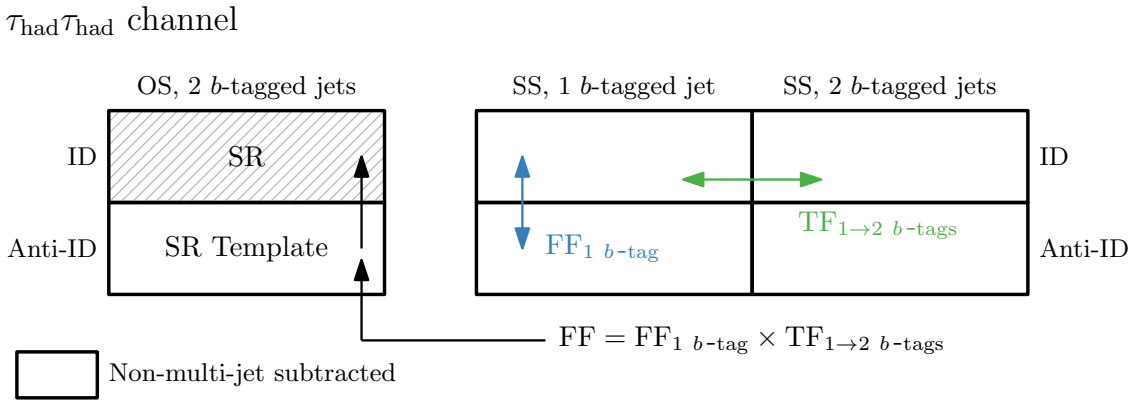
Figure 3.3: Schematic depiction of the combined fake-factor method used to estimate multi-jet and $t\bar{t}$ backgrounds with fake-$\tau_{\text{had-vis}}$ in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. Backgrounds which are not from events with fake-$\tau_{\text{had-vis}}$ originating from jets are estimated from simulation and are subtracted from data in all control regions. Events in which an electron or a muon is misidentified as a $\tau_{\text{had-vis}}$ are also subtracted, but their contribution is very small. Both sources are indicated by 'True-$\tau_{\text{had-vis}}$ subtracted' in the legend.

separately. The former is measured with a fake-factor method as depicted in Figure 3.4(a) while the latter is measured with a scale-factor method as depicted in Figure 3.4(b).

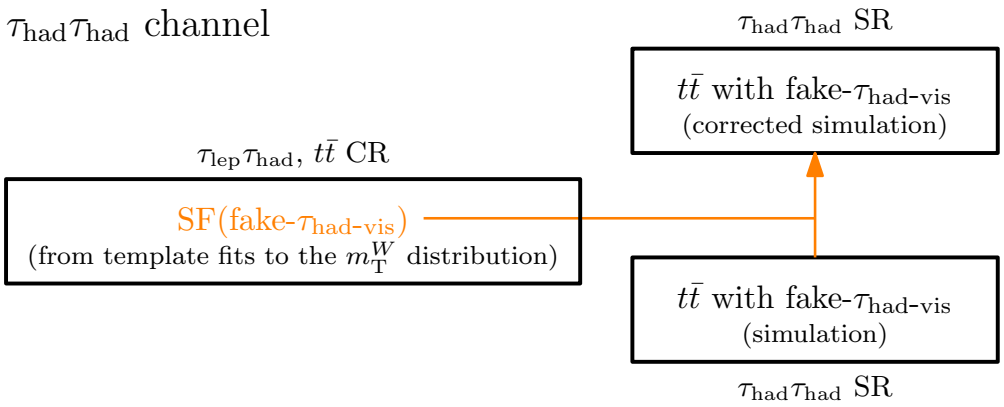
3.3.2 Strategy validation

Since the fake-$\tau_{\text{had-vis}}$ background estimation strategy is inherited from previous round of the analysis for both $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels, it should be validated, especially the background estimation for $\tau_{\text{had}}\tau_{\text{had}}$ channel where the previously-derived fake factors and scale factors are re-used in this round of the analysis. For the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, new fake factors are re-derived and the strategy is validated in the dedicated 0–b–tag region. In the following, the validation of the background estimation strategy for $\tau_{\text{had}}\tau_{\text{had}}$ channel is presented. It's found that the strategy is still feasible and the background estimation can be accurately performed even with the modification of the Monte Carlo configuration used to model the V+jets samples from Sherpa 2.2.1 to Sherpa 2.2.11, and the inclusion of $t\bar{t}$ dilepton samples.

The validation of fake-$\tau_{\text{had-vis}}$ background estimation in $\tau_{\text{had}}\tau_{\text{had}}$ channel includes two parts, one is to validate the estimation of multi-jet with fake-$\tau_{\text{had-vis}}$ while the other is to validate the estimation of $t\bar{t}$ with fake-$\tau_{\text{had-vis}}$. The idea is to check the data/mc consistency in dedicated validation regions. For example, in terms of the multi-jet with fake-$\tau_{\text{had-vis}}$, it's checked in the validation regions like 1–tag SS region, 1–tag OS region; in terms of the $t\bar{t}$ with fake-$\tau_{\text{had-vis}}$, it's checked in the dedicated Top validation region where the true $t\bar{t}$ estimation is also validated; there is also a dedicated validation region called 2–tag SS region where the fake-$\tau_{\text{had-vis}}$ originating from jets in multi-jet and $t\bar{t}$ is validated.



(a) The fake-factor method to estimate the multi-jet background with fake-$\tau_{\text{had-vis}}$. Backgrounds that are not from multi-jet events are simulated and subtracted from data in all the control regions. This is indicated by 'Non-multi-jet subtracted' in the legend



(b) The scale-factor method to estimate the $t\bar{t}$ background with fake-$\tau_{\text{had-vis}}$. The scale factor is firstly calculated in the $\tau_{\text{lep}}\tau_{\text{had}}$ $t\bar{t}$ CR, then applied to the simulated $t\bar{t}$ with fake-$\tau_{\text{had-vis}}$ in the $\tau_{\text{had}}\tau_{\text{had}}$ SR

Figure 3.4: Schematic depiction of the background estimation strategy for both the multi-jet and $t\bar{t}$ backgrounds with fake-τ_{had} in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

From the following validation plots, it's concluded that the fake-$\tau_{\text{had-vis}}$ background estimation strategy is still valid and has good performance.

3.3.2.1 Fake factor method validation : 1–tag SS

If the reversal of the Tau-ID would affect the shape of event variables other than variables that the fake factors are binned in, then this would be visible as non-closure in the 1-tag SS ID-region (the numerator of the fake factor calculation). In Figures 3.5 to 3.7 this closure check is shown in the distributions of a representative set of variables. This closure test is updated concerning the previous round of the analysis as it is performed with the V+jets SHERPA 2.2.11 and $t\bar{t}$ dilepton samples. Previously, a significant non-closure was observed between fake τ_{had} in the barrel and endcap. This was largely resolved by introducing an endcap bin in τ_{had} η for the fake factor estimation. However, some residual shape within the endcap is observed (this can only be resolved by even finer τ_{had} η binning of the fake factors) but will have a negligible effect on the 2-tag SR.

This region is used to estimate the fake factors and can therefore only contribute as a closure check.

3.3.2.2 Fake factor method validation : 1–tag OS

The validation of the multi-jet fake τ_{had} estimate is shown in Figures 3.8 to 3.10. The OS multi-jet validation region is defined as follows:

$$N_{\text{b-tag}} = 1, \qquad \text{OS } \tau_{\text{had}} \text{ charge}, \qquad m_{\text{MMC}} > 110\,\text{GeV}, \qquad \frac{\text{MET}}{\sigma_{\text{MET}}} < 3$$

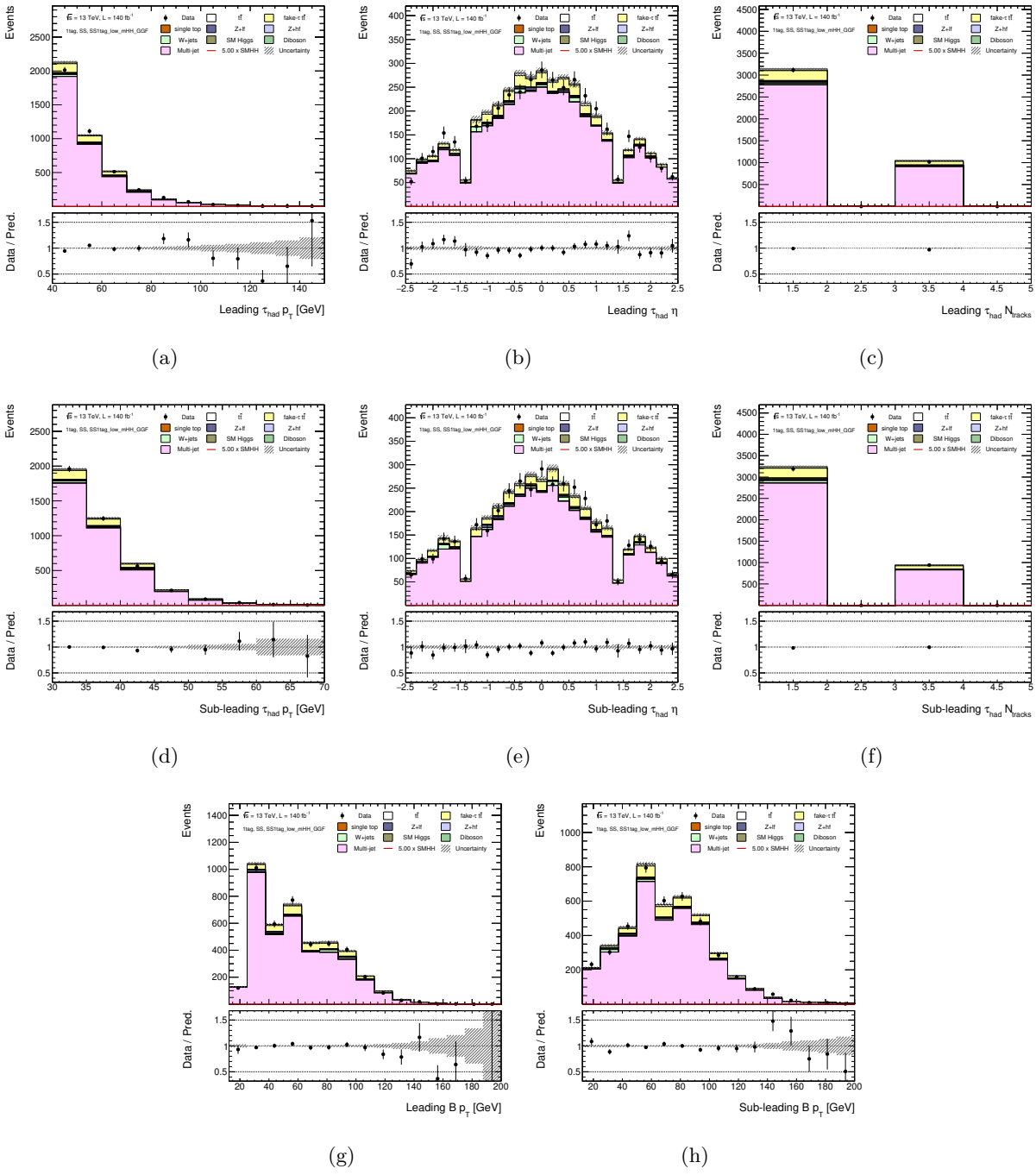
This region is not used in the development of the method and therefore constitutes an independent validation region.

3.3.2.3 Top validation region

The top validation region is defined as:

$$N_{\text{b-tag}} = 2, \qquad \text{OS } \tau_{\text{had}} \text{ charge}, \qquad m_{\text{MMC}} + m_{\text{bb}} > 300\,\text{GeV}, \qquad m_{\text{MMC}} > 110\,\text{GeV}$$

and therefore overlaps partially with the signal region. However, due to the $m_{\text{MMC}} + m_{\text{bb}}$-cut the signal contamination in this region is negligible. Pre-fit plots of this region are shown in Figures 3.11 to 3.13 and the same $t\bar{t}$ mis-modelling observed in the previous round of the analysis [79] is found.



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 3.5: Closure check of the fake factor method in 1-tag SS ID-region for the low-m_{HH} ggF category in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

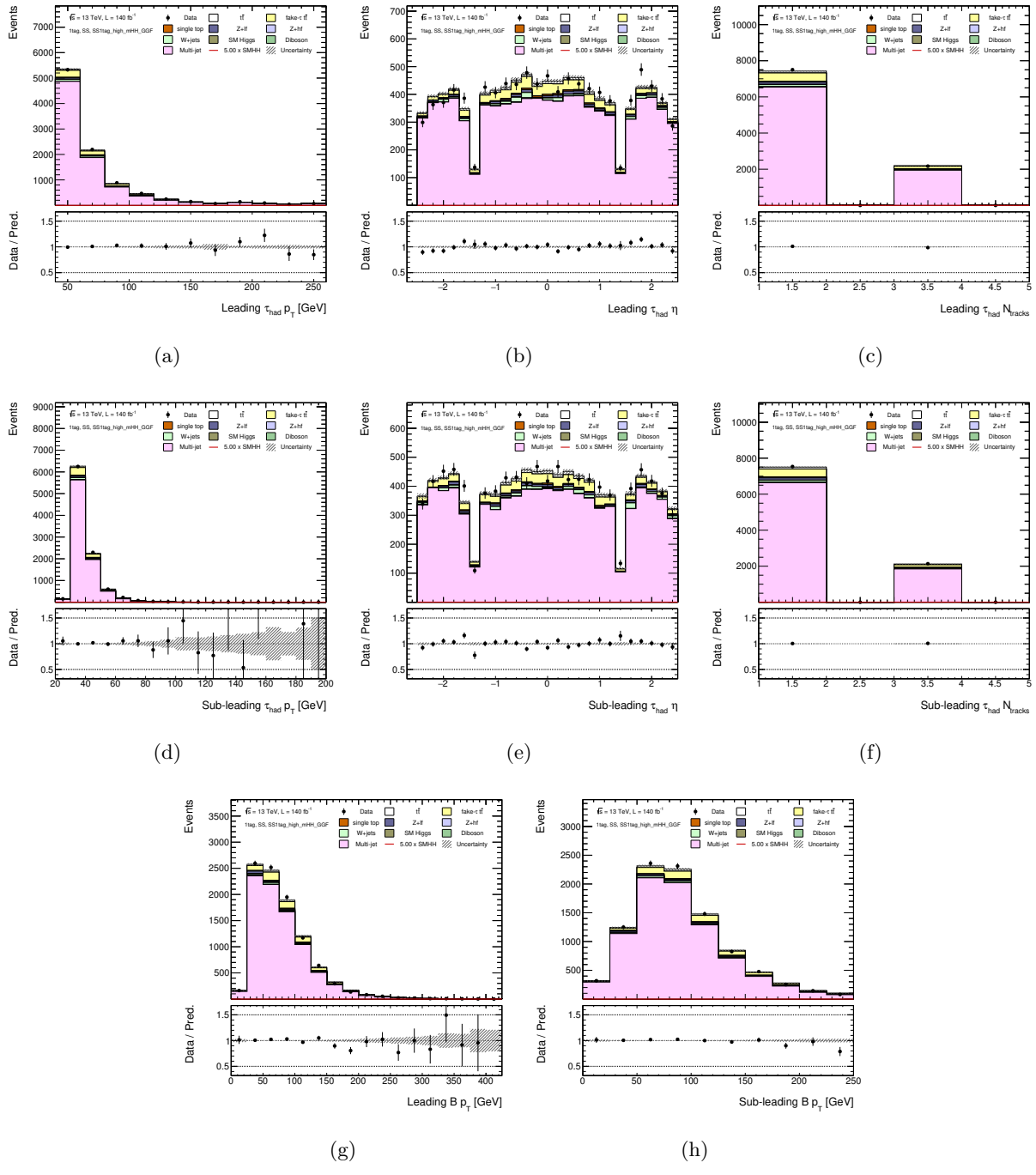
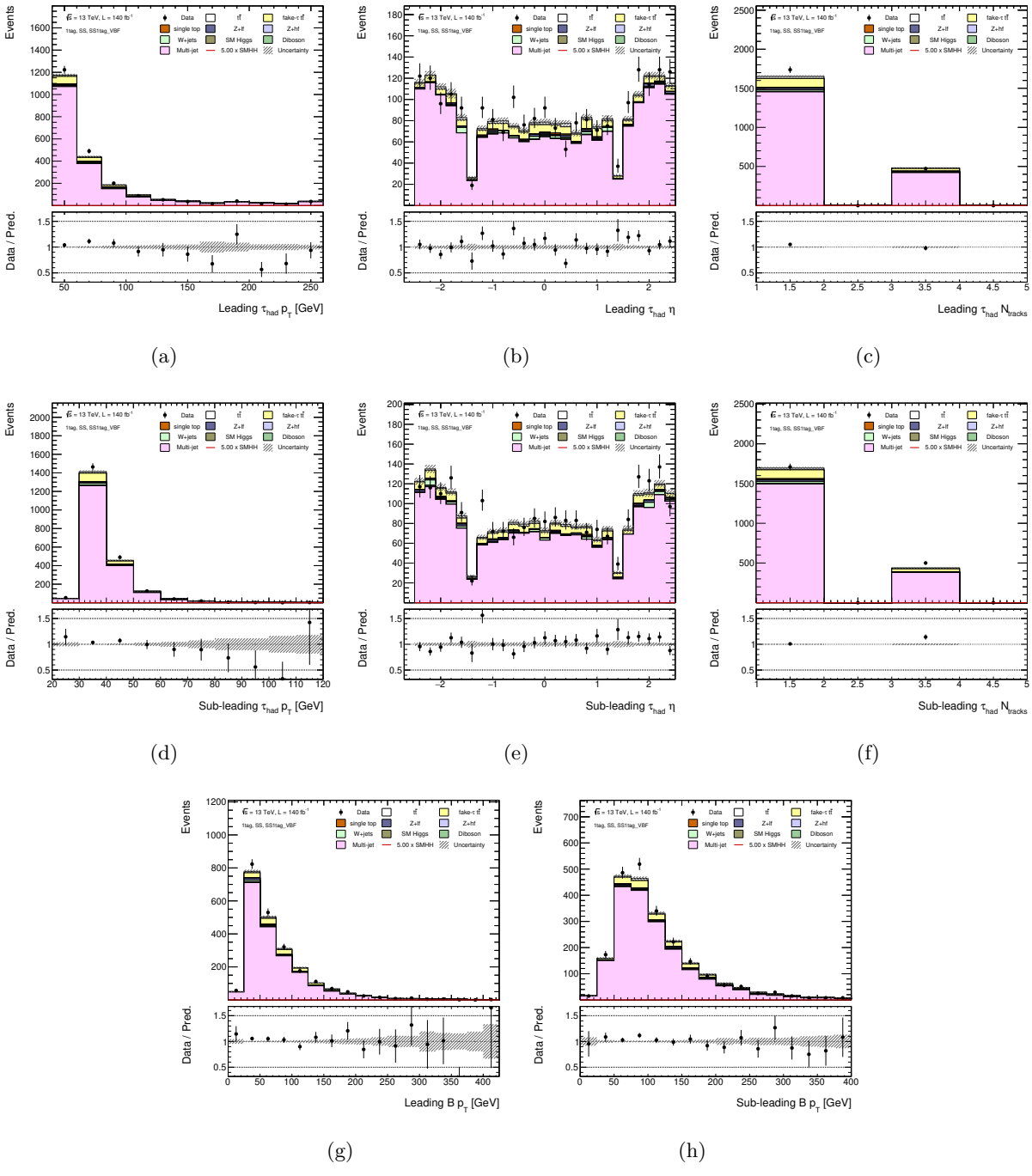


Figure 3.6: Closure check of the fake factor method in 1-tag SS ID-region for the high-m_{HH} ggF category in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 3.7: Closure check of the fake factor method in 1-tag SS ID-region for the VBF category in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

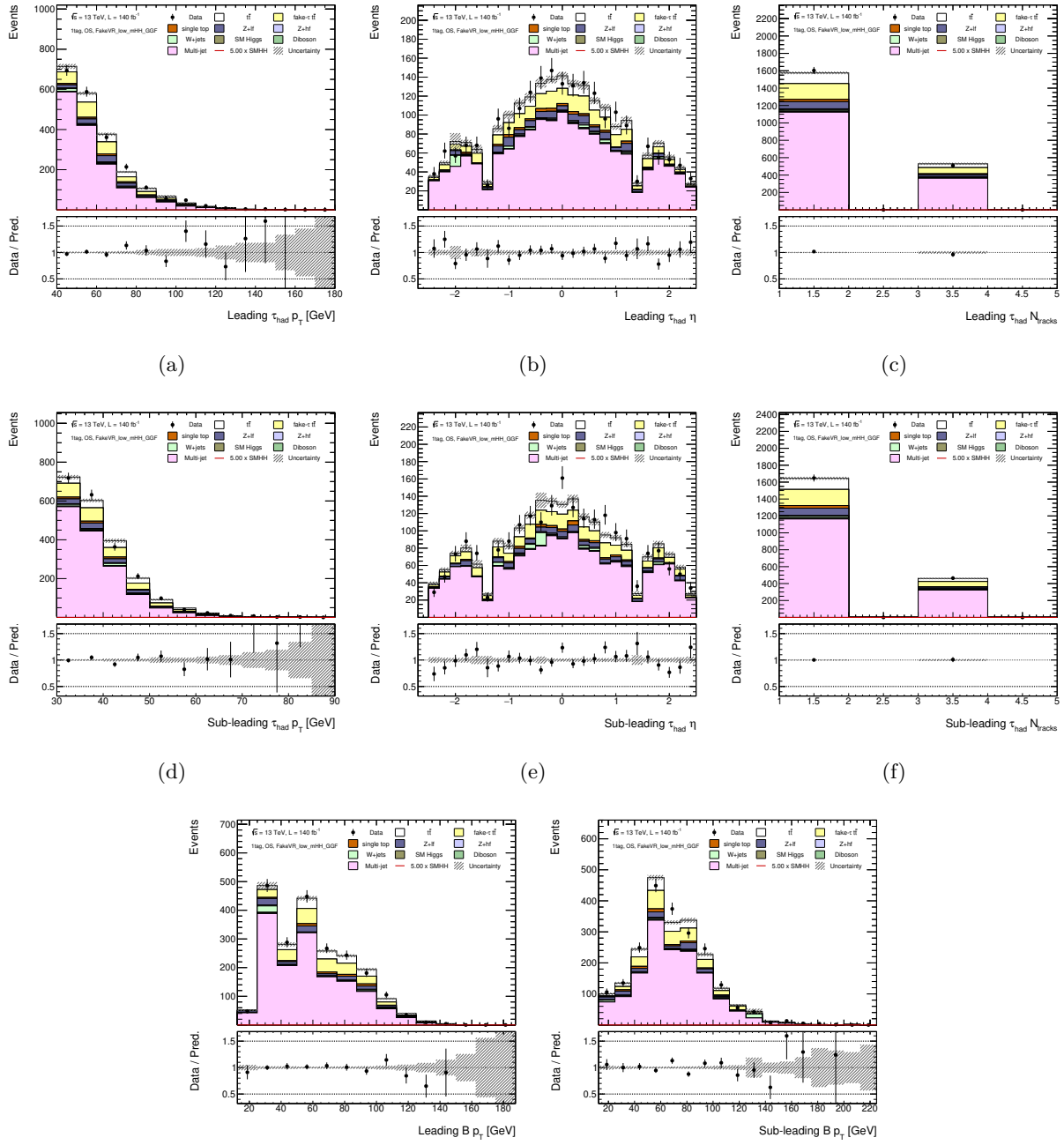
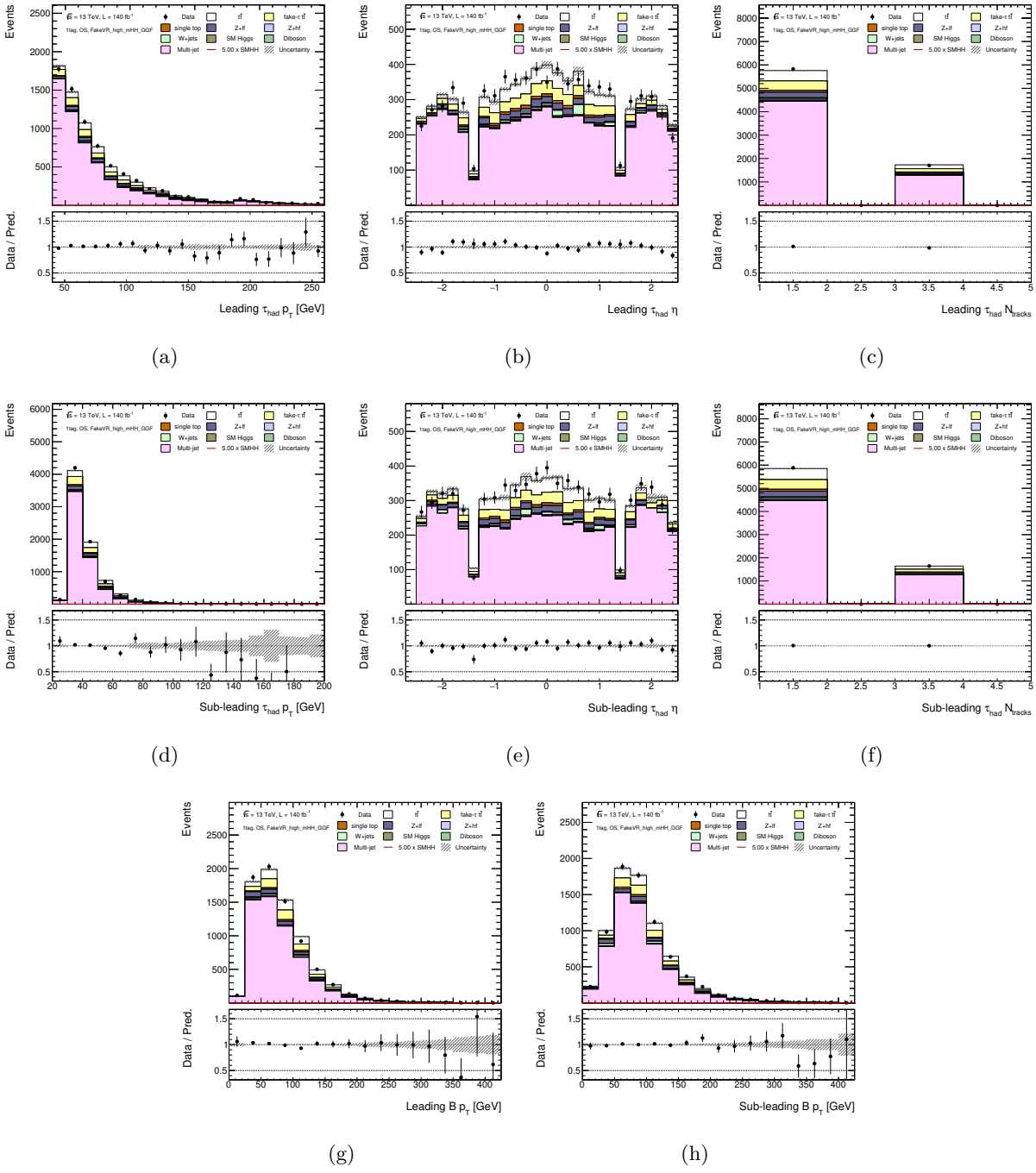


Figure 3.8: Validation of the multi-jet estimate in the 1-tag OS multi-jet validation region for the low-m_{HH} ggF category of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Uncertainties are from statistical sources only. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

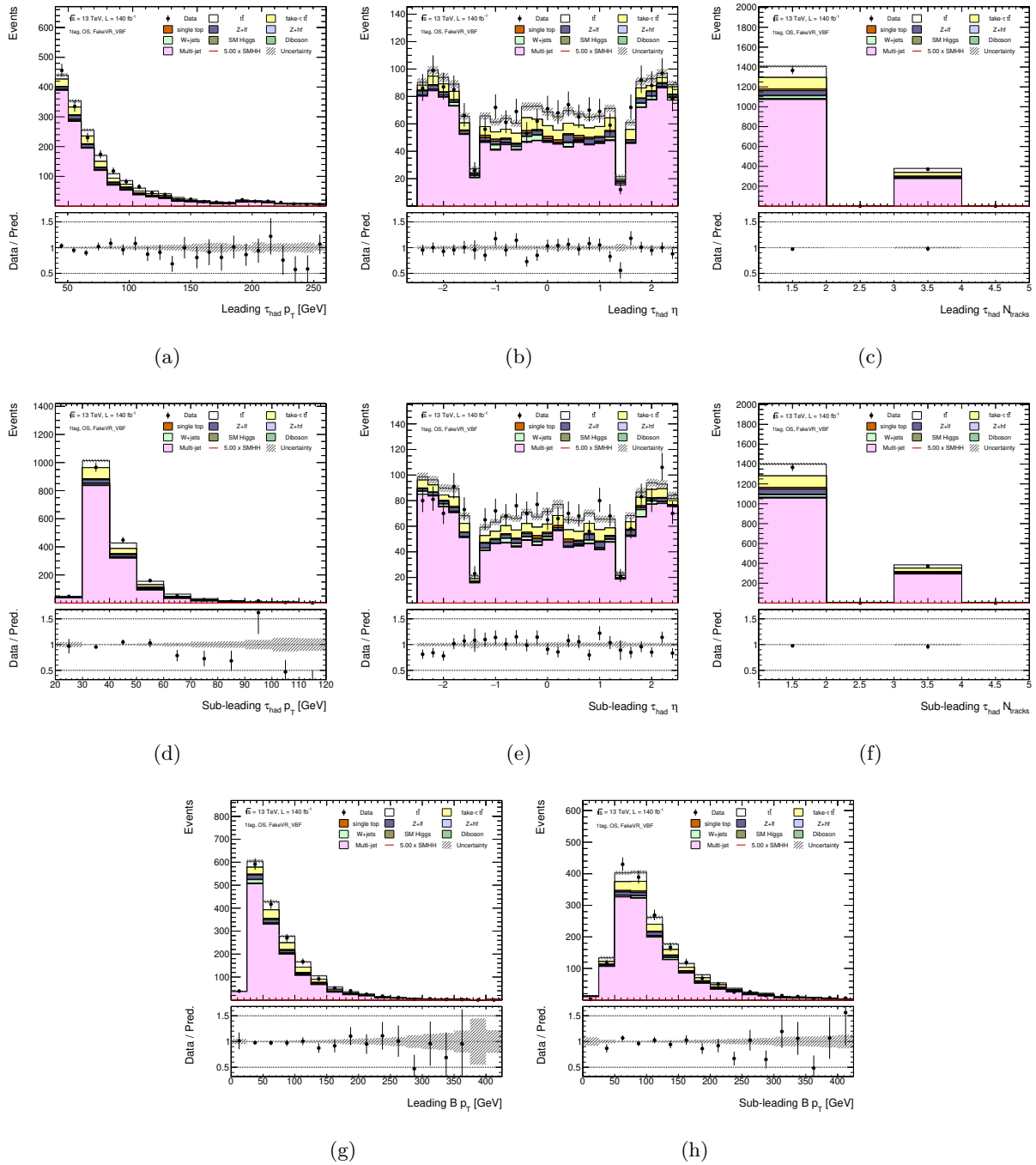


(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 3.9: Validation of the multi-jet estimate in the 1-tag OS multi-jet validation region for the high-m_{HH} ggF category of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Uncertainties are from statistical sources only. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 3.10: Validation of the multi-jet estimate in the 1-tag OS multi-jet validation region for the VBF category of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Uncertainties are from statistical sources only. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

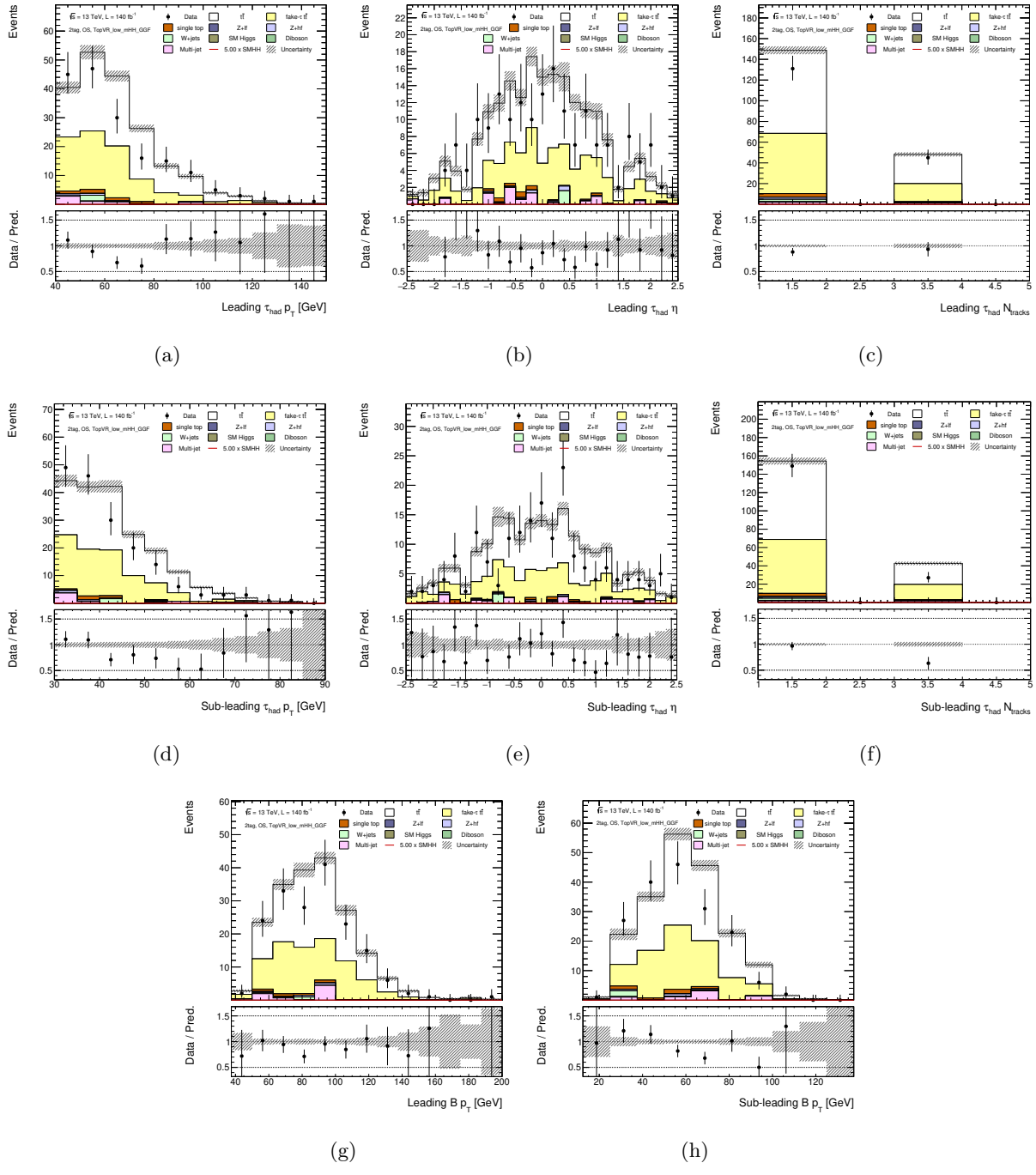


Figure 3.11: Validation of the $t\bar{t}$ estimate in the 2-tag OS top validation region for the low-m_{HH} ggF category in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

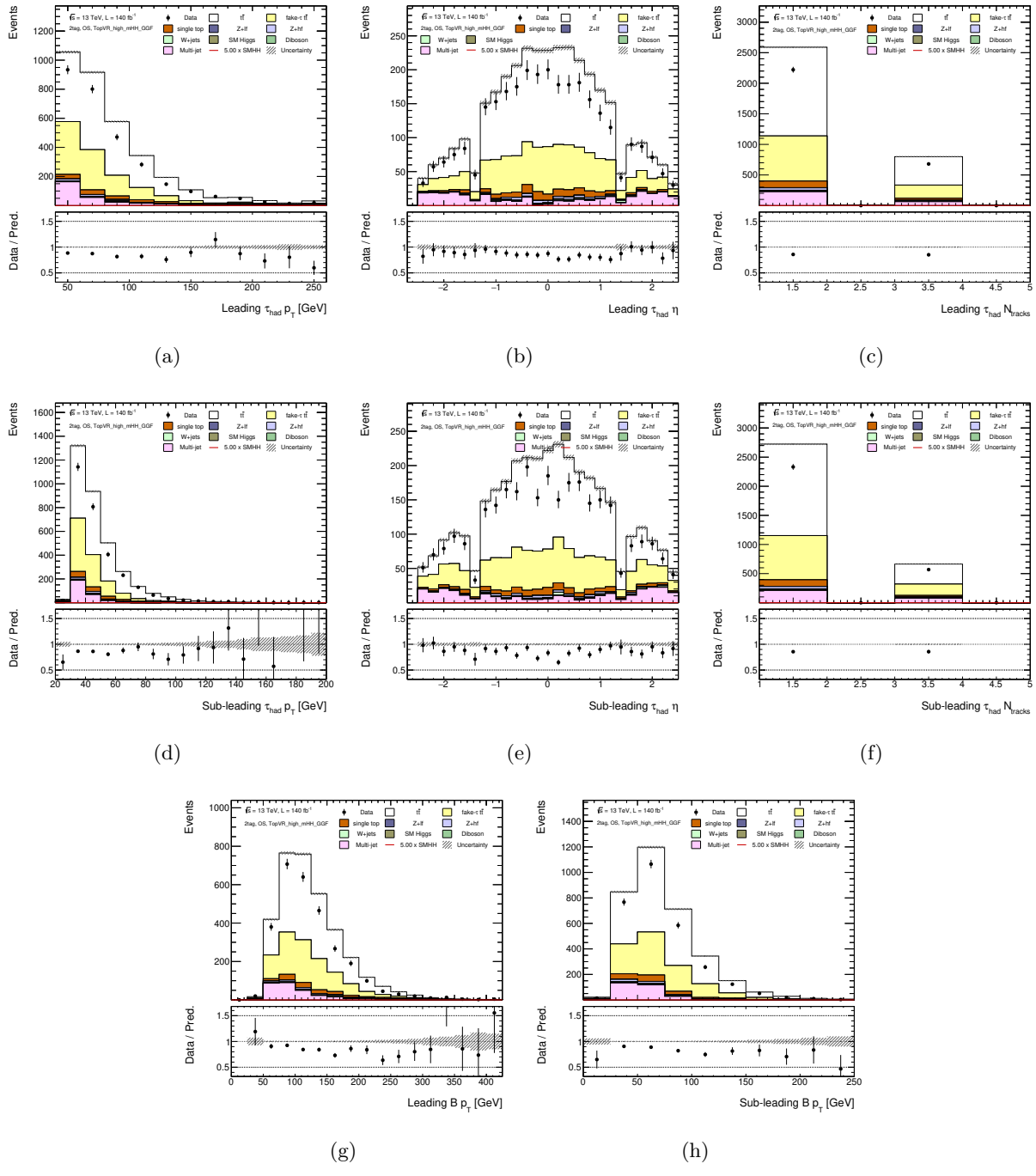


Figure 3.12: Validation of the $t\bar{t}$ estimate in the 2-tag OS top validation region for the high-m_{HH} ggF category in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

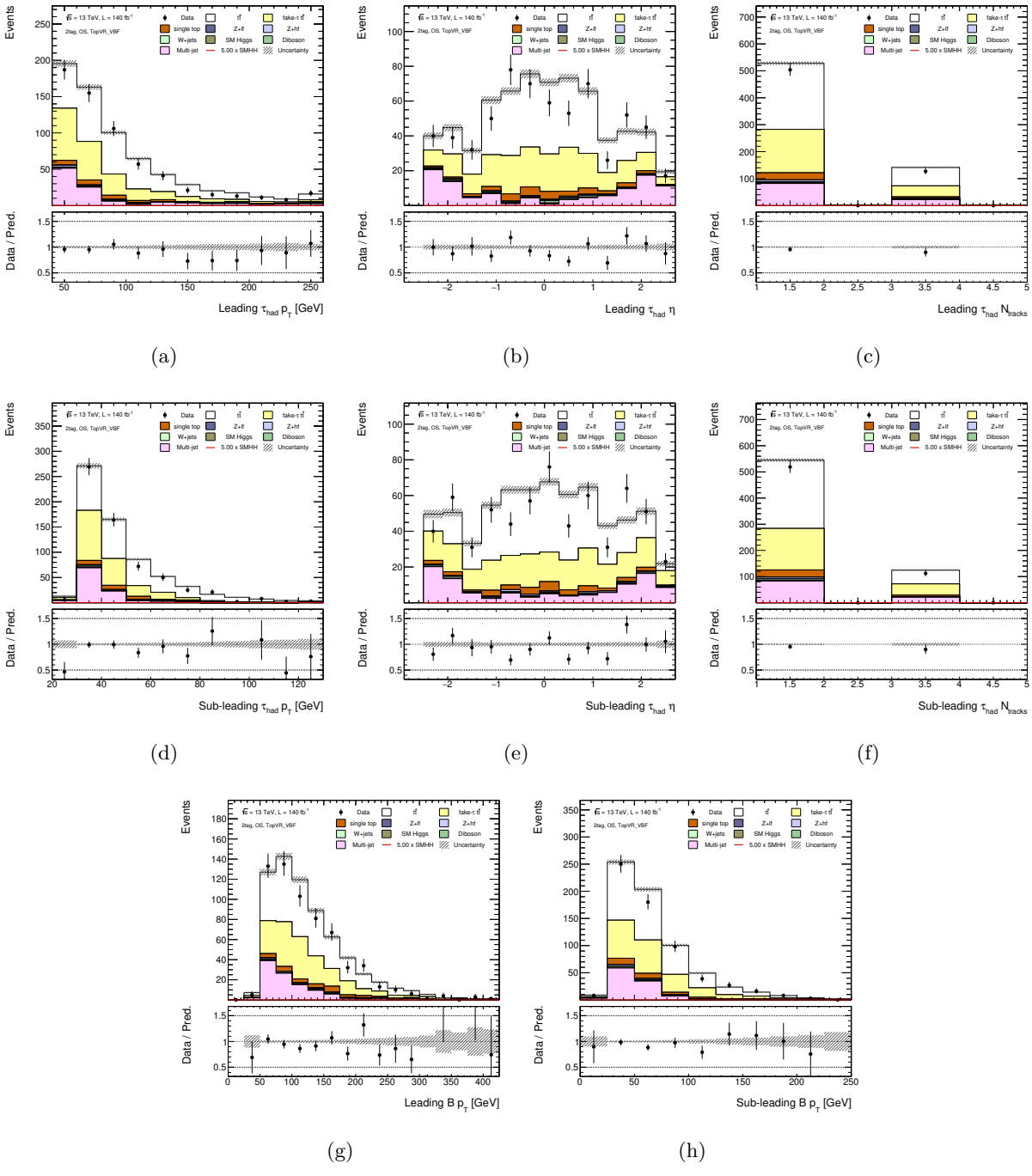


Figure 3.13: Validation of the $t\bar{t}$ estimate in the 2-tag OS top validation region for the VBF category in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

3.3.2.4 Fake validation region

The 2-tag SS region after preselection is enhanced in τ_{had} originating from jets in multi-jet and $t\bar{t}$. Prefit plots of this region are shown in Figures 3.14 to 3.16.

3.4 Multivariate signal extraction

The output of a multivariate algorithm is used as a final discriminant for the signal extraction. In contrast to the previous full Run-2 analysis round, BDTs are used in both sub-channels across all categories. A larger number of input variables have been explored, and training hyper-parameters chosen to maximize the model performance. The same training approach is used across both $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels, and a common training framework built on TMVA [123], is employed in all cases.

3.4.1 General MVA and optimization strategy

3.4.1.1 Folding strategy

The design of the MVA strategy is driven by the requirement of ensuring a reliable and unbiased estimate of the expected analysis sensitivity. This implies that simulated events may not simultaneously be used in the definition of the BDT (including its hyper-parameters and input variables) and in the generation of the histogram templates for the BDT output score distributions.

The simplest strategy compatible with this requirement, implemented here, is to partition the available set of simulated events into three partitions of equal size. In our concrete implementation, the partitioning is performed based on the event number.

Then, three separate BDT models are trained, each using a separate partition of the available simulated events ("training folds") as shown in Table 3.2. The same hyper-parameters and input variables are used for each training. They are selected by optimizing the performance of the BDT achieved on the "validation folds", not seen during training.

Table 3.2: Partitioning of the available sets of simulated events for the training, optimization, and evaluation of the BDT models.

Model	Fold 0 `event_number %3 = 0`	Fold 1 (`event_number %3 = 1`)	Fold 2 (`event_number %3 = 2`)
BDT 0	Training	Validation	Testing
BDT 1	Testing	Training	Validation
BDT 2	Validation	Testing	Training

The simulated events contained in the "testing folds" are used in the creation of the histogram fit templates. By construction, these events do not enter the BDT optimization

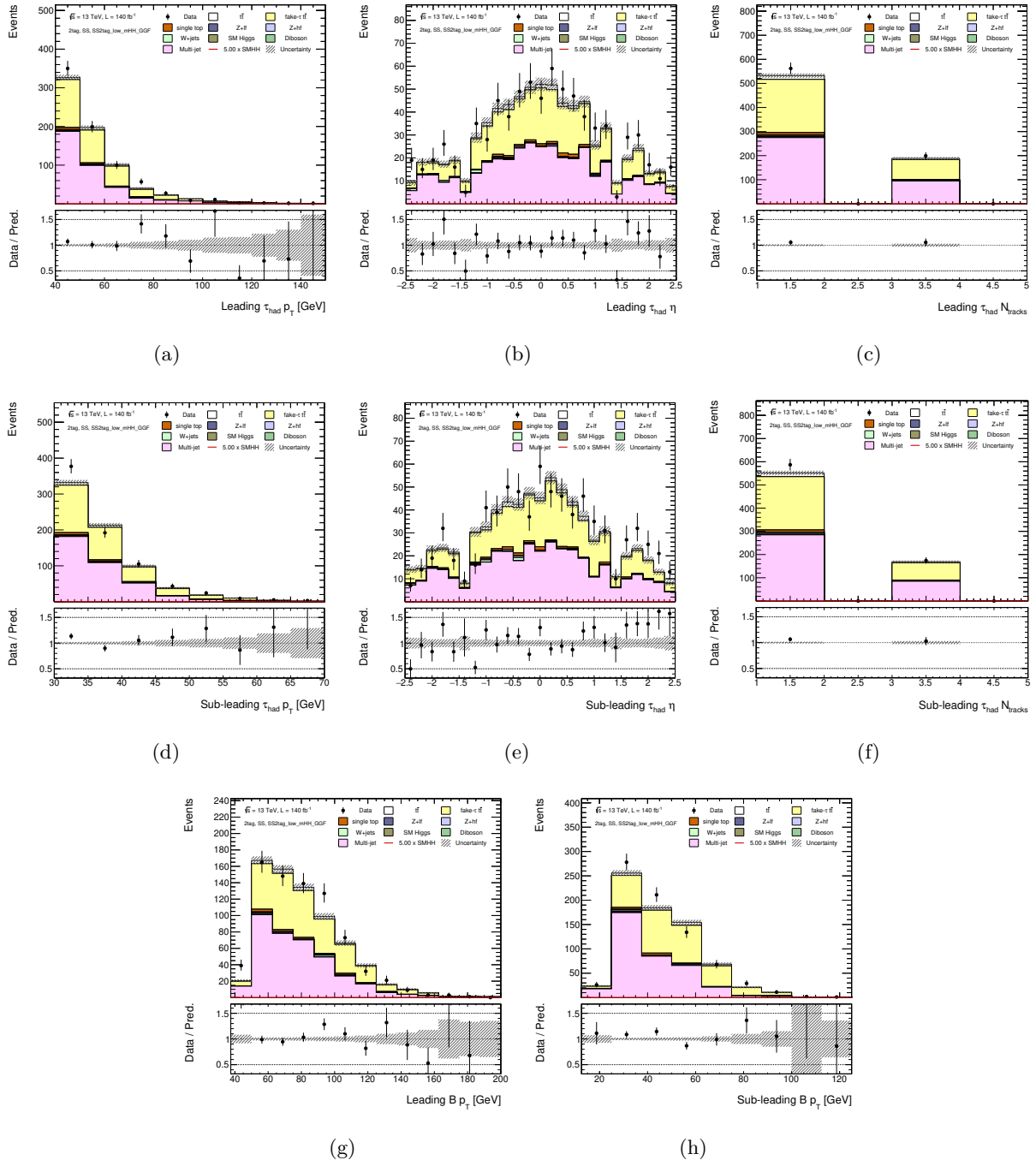


Figure 3.14: Validation of the $t\bar{t}$ estimate in the 2-tag SS fake validation region for the low-m_{HH} ggF category in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

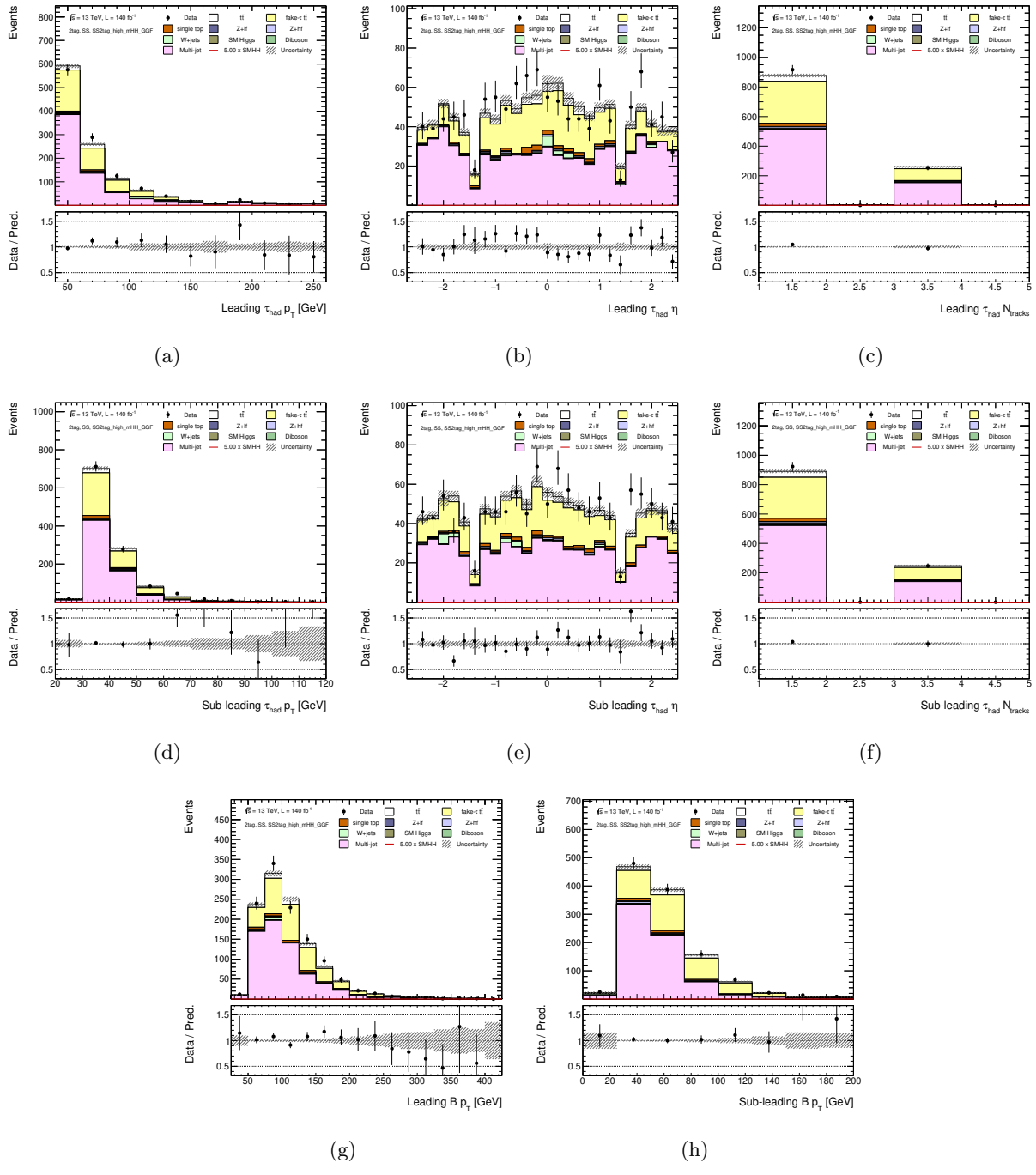


Figure 3.15: Validation of the $t\bar{t}$ estimate in the 2-tag SS fake validation region for the high-m_{HH} ggF category in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

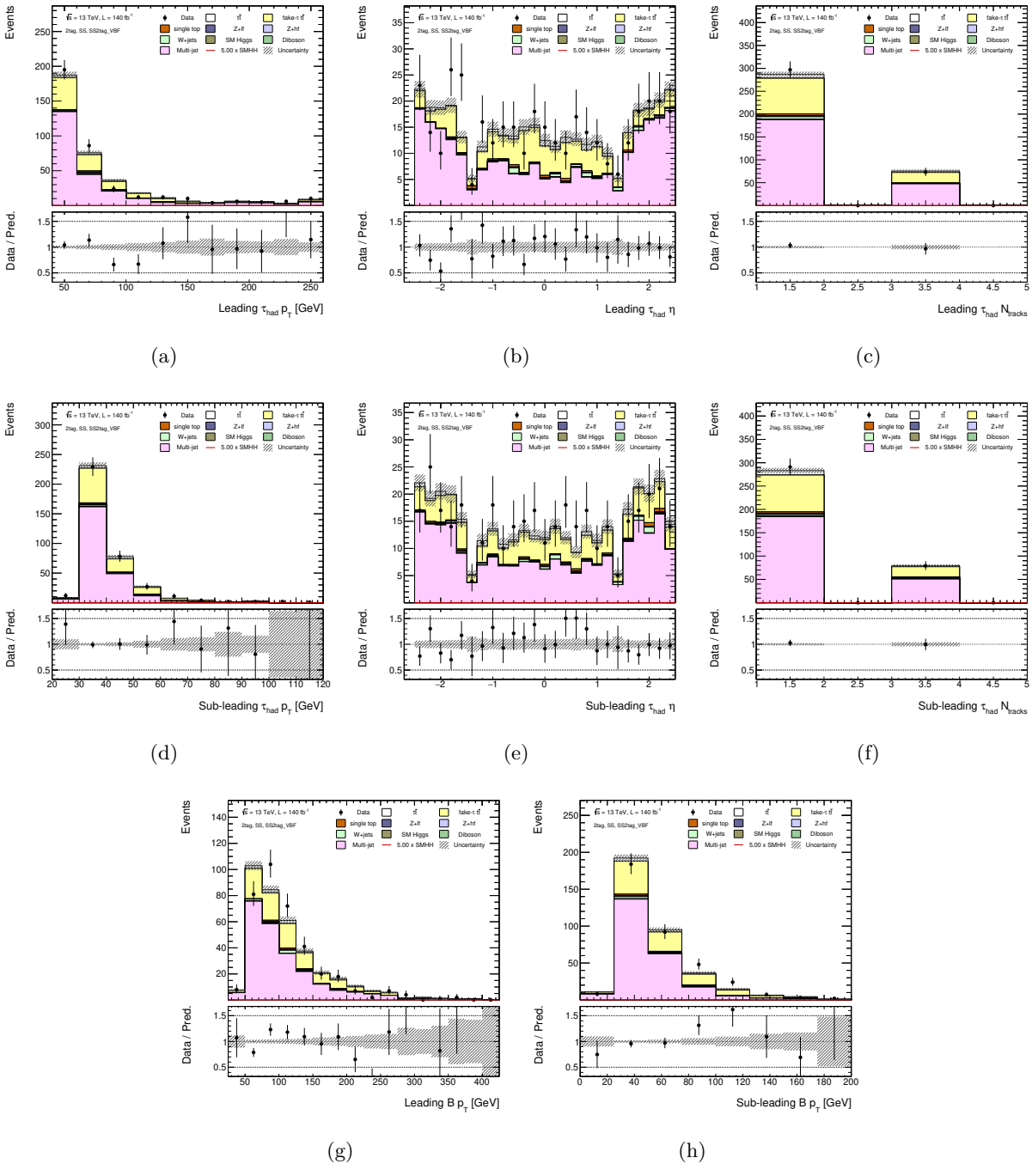


Figure 3.16: Validation of the $t\bar{t}$ estimate in the 2-tag SS fake validation region for the VBF category in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Shown are the leading and sub-leading τ_{had} p_{T} (a,d), η (b,e), N_{track} (c, f), and b-jet candidate p_{T} after energy correction (g,h).

procedure, ensuring an unbiased estimate of the expected analysis significance.

3.4.1.2 Optimization of hyper-parameters

Certain TMVA training hyper-parameters may significantly impact the classification accuracy of the BDT. Of particular importance are the number of trees (`NTrees`) in the ensemble, as well as the depth of each decision tree (`MaxDepth`).

A joint optimization is performed over these two hyper-parameters. For a particular set of hyperparameter values, a set of BDTs is trained according to the folding procedure outlined in Section 3.4.1.1. The binned distribution of the BDT output score is then computed separately for signal and the sum of all backgrounds, using the simulated events contained in the validation folds. (The binning is computed by the same algorithm that also defines the binning used in the likelihood fit, as described in Section 3.6.) The binned signal significance is used as a performance metric,

$$Z = \sqrt{\sum_{i \in \text{bins}} 2\left((s_i + b_i)\log\left(1 + \frac{s_i}{b_i}\right) - s_i\right)}, \tag{3.1}$$

where s_i and b_i are the expected numbers of signal and background events in bin i, respectively.

Hyperparameter values are investigated within a predefined range, depending on the analysis region and type of BDT. Bayesian optimization is used to preferentially select hyperparameter settings resulting in higher values for the binned significance.

The final set of hyper-parameters is selected from the list of parameter points seen during optimization.

3.4.1.3 Selection of input variables

Many kinematic variables with discriminating power may be defined for the $bb\tau\tau$ system. However, it is desirable to select a "minimal" set of inputs by removing highly correlated or redundant observables, which nevertheless is "complete" and enables near-maximal discrimination power. (Note that this selection of variables is not strictly necessary—provided that all variables used as inputs to the BDT are well-modelled—but rather viewed as a simplification.)

Starting from the list of all available variables, a reasonably minimal set is constructed iteratively:

- A small set of "baseline" input variables is always used. (The particular variables in the baseline set depends on the BDT, and is explained below.)
- Additional variables are selected from the remaining list in a greedy fashion: at each step, the observable leading to the greatest improvement of the binned signal signifi-

cance (computed according to Eq. 3.1 on the validation folds) is selected.

- If no additional variable leads to an improvement in the validation significance, the variable leading to the smallest loss is added instead.
- If no improvement is seen during N_p consecutive steps, the optimization procedure is stopped.

Such greedy optimization procedures are highly susceptible to the presence of statistical noise in the performance measure used to steer the algorithm. To reduce the level of statistical fluctuations in the significance measure Z to a level smaller than the performance gain expected from the selection of input variables, the binned significance is typically computed on a coarser binning than the one used in the fit, as explained below.

3.4.2 ggF/VBF BDT

The ggF/VBF classification BDT is trained to discriminate between the SM gluon-fusion HH and the SM VBF HH signal processes (see Section 3.2.4). The aim of this classification is to improve the κ_{2V} constraint by creating a dedicated category for VBF events. The BDT is only applied to VBF candidate events, and only such events are used in the training. (VBF candidate events are defined as events that have at least two jets in addition to the $H \longrightarrow bb$ ones. Nearly half of the ggF HH events pass the VBF candidate requirement and the other half fail, while around 80% of the VBF HH events pass.) The BDT is trained so that ggF-like events are assigned with the scores near 1.0 while VBF-like events are assigned with the scores near -1.0. Accordingly, events with the scores below a chosen cut value (working point) fall into the VBF category.

The ggF/VBF classification BDT adopts the strategy described in Section 3.4.1 to perform the 3-fold training, input variable selection and hyperparameter optimization. For the working point selection, it is defined as the optimal cut value on the output of the ggF/VBF classification BDT that gives the best analysis sensitivity.

3.4.3 SR BDT

In each of the nine SRs as shown in Figure 3.1, a dedicated set of BDTs is trained to discriminate the HH signal against the sum of all SM backgrounds.

In terms of the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the low-m_{HH} region BDT is trained on the ggF $\kappa_\lambda = 10$ signal events, while the ggF SM signal sample is used in the high-m_{HH} region. In both cases, the training is performed on events falling into the respective SR, i.e. the full event selection, including the cut on the ggF/VBF BDT, is applied. The signal-vs-background BDT in the VBF region is trained on SM VBF HH signal events. In this case, events in both the VBF and the ggF SRs are used in the training to maximize the available statistics.

This was found to result in a significant improvement in the VBF sensitivity compared to the situation where only events in the VBF SR are used in the training.

In terms of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channels, the low-m_{HH} region BDT is trained on the ggF $\kappa_\lambda = 10$ signal events, while the ggF SM signal sample is used in the high-m_{HH} region. The VBF BDT classifier, on the other hand, was trained specifically to distinguish VBF HH events from combined background events in the VBF region.

The hyper-parameters for the BDTs as well as their input variables are chosen separately for each region, as explained below. The obtained BDT scores are then fitted using the profile likelihood fit to obtain the final results, which are discussed in Section 3.7.

3.5 Systematic uncertainties

Systematic uncertainties related to the background and signal estimation in the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ signal regions and Z+HF control region are evaluated and propagated to the final statistical analysis fit. These include both experimental and theoretical modelling uncertainties, as summarized below. They are expressed as uncertainties in the overall yield and in the shape of the final observable, i.e. the MVA score distributions for each analysis category. Given the new categorization and MVAs in this analysis round, theoretical modelling uncertainties have to be re-derived.

3.5.1 Experimental uncertainties

Experimental uncertainties[2] include those stemming from the measured integrated luminosity, the modelling of the pileup, as well as those related to the physics object reconstruction and identification. They are estimated by tools provided by the ATLAS combined performance (CP) groups. These uncertainties are applied to all MC-based processes.

Different reduction schemes are available for the systematic uncertainties related to the jet energy resolution (JER) and jet energy scale (JES), b-tagging and e-gamma energy scale. The CategoryReduction and FullJER scheme are used to derive uncertainties affecting JES and JER. This is the recommended set of systematic uncertainties for analyses meant to be combined with other searches, and it is therefore adopted by all HH analyses. For the b-tagging uncertainties, the medium reduction scheme is used by all HH analyses. It contains 13 NPs; 3 NPs for b-jets, 4 NPs for c-jets, 4 NPs for light-jets and 2 NPs for extrapolations. The simplified correlation model is used for the e-gamma uncertainties, as suggested for analyses which are very weakly sensitive to the energy scale of electrons and photons.

[2]This part hasn't changed with respect to the previous full Run-2 analysis.

3.5.2 Modelling uncertainties on MC-based background

The background modelling uncertainties on MC-based processes are comprised of theoretical cross-section and acceptance uncertainties. Cross-section uncertainties only affect normalizations and are applied to all backgrounds[3] estimated using simulation, following the Physics Modelling Group (PMG) recommendations. Acceptance uncertainties can affect both normalizations and shapes. Their contribution is usually split into a normalization acceptance uncertainty affecting the signal region event yields and a shape uncertainty affecting the shape of the final discriminant distributions.

The acceptance uncertainties applied on minor backgrounds (Z+ light-flavor jets, W+ jets and Diboson) are taken from the SM $VHbb$ analysis. Acceptance uncertainties for $t\bar{t}$, Z+HF jets, single-top (Wt channel), $t\bar{t}H$ and ZH processes are estimated in the signal regions – and control regions for $t\bar{t}$, Z+HF jets – of this analysis from comparisons between nominal and alternative MC simulated samples or samples with varied event weights.

The normalization acceptance uncertainties are derived by comparing the acceptance, A, obtained for each alternative sample i to the nominal sample acceptance. Equivalently, we compare the number of expected events, N, for each variation i to the nominal expected events in the signal regions for all samples that are normalized to the same cross-section:

$$\sigma_A^i = \frac{A^i_{\text{variation}} - A_{\text{nominal}}}{A_{\text{nominal}}} = \frac{N^i_{\text{variation}} - N_{\text{nominal}}}{N_{\text{nominal}}}\,. \tag{3.2}$$

If one of the regions included in the final fit is better than the others at constraining some normalization factor (NF), a relative normalization acceptance uncertainty is assigned. It is estimated from the comparison of the relative amount of events predicted by the alternative model in one region R with respect to another region R' and the same fraction in the nominal model:

$$\sigma_{A_R/A_{R'}} = \frac{\frac{A_{\text{variation},R}}{A_{\text{variation},R'}} - \frac{A_{\text{nominal},R}}{A_{\text{nominal},R'}}}{\frac{A_{\text{nominal},R}}{A_{\text{nominal},R'}}} = \frac{\frac{N_{\text{variation},R}}{N_{\text{variation},R'}} - \frac{N_{\text{nominal},R}}{N_{\text{nominal},R'}}}{\frac{N_{\text{nominal},R}}{N_{\text{nominal},R'}}}\,. \tag{3.3}$$

Acceptance uncertainties on the normalization of backgrounds are applied in all regions and treated as correlated across regions, except for the $t\bar{t}$ and Z+HF backgrounds, whose normalizations are left unconstrained (floated) in the global likelihood fit. For those two processes, relative acceptance uncertainties between regions with a common underlying normalization are applied.

Shape acceptance uncertainties are applied when differences are identified in the final discriminant distributions of the nominal and alternative samples. The comparison is done by taking the normalized distributions of the BDT scores, and calculating the ratio of the

[3]Except for $t\bar{t}$ and Z+HF processes which have freely floating normalization factors (NFs) in the fit.

alternative and nominal number of events per bin.

The parameterization of shape uncertainties given in terms of the BDT score distribution is not tied to the binning used in the statistical analysis. A dedicated re-binning algorithm, summarized below, is used to ensure that only statistically significant shape effects are retained. The algorithm depends on three parameters t_f, t_s and t which are used to define two requirements on the merged bins. The first condition is that the relative MC statistical uncertainty on the merged bin content is below t_s for each of the two compared distributions, and it is applied for bins in the signal-like BDT score regime. The second condition is for the fraction of total events in the merged bin to exceed t_f for each of the two compared distributions. This condition is applied in the background-like regime. The algorithm starts merging pre-bins from the most signal-like BDT scores, and once the BDT score value t is reached, the deciding merging criterion is changed from the first to the second.

The values of t_f, t_s, and t depend on the variation and the analysis region. They are manually chosen to ensure a healthy and smooth parameterization for each uncertainty component.

In the following, the uncertainties on single-top for $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel is presented as an example. For the other uncertainties, they are treated in a similar manner thus not described.

3.5.2.1 Uncertainties on single-top in lephad SLT channel

Single-top acceptance (normalization + shape) uncertainties are derived by MC-to-MC comparisons, following the Top Modelling Group recommendations. The PS, ME and $t\bar{t}$ interference (Wt-channel) uncertainties are estimated by the differences between the nominal and corresponding alternative samples listed in Table 3.3. Only the contribution from the Wt-channel is considered since it dominates over the s- and t-channel contributions. For PS uncertainties, AF2 nominal samples showered with Pythia8 are compared against alternative AF2 samples showered with Herwig7, while the ME uncertainty is evaluated by comparing the AF2 nominal Powheg+Pythia8 samples to the alternative AF2 MadGraph5_aMC@NLO+Pythia8 samples. The detector acceptance uncertainty of single-top production related to the interference between the Wt and $t\bar{t}$ processes is estimated via comparison of the full-simulation nominal Powheg+Pythia8 samples with the diagram removal (DR) scheme to the alternative full-simulation Powheg+Pythia8 samples with the diagram subtraction (DS) scheme. Uncertainties due to PDF, ISR and FSR are evaluated using internal alternative weights present in the single-top nominal samples.

Furthermore cross section uncertainties are applied following the recommended NNLO calculations; the values are reported in Table 3.4 for the three production modes.

The normalization acceptance uncertainties obtained from the described MC-to-MC comparisons in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel are summarized in Table 3.5.

Table 3.3: List of alternative single-top (Wt) samples.

Source	DSID	Name
ME	412002	aMcAtNloPythia8EvtGen_HThalfscale_tW_inclusive
PS	411036, 411037	PowhegHerwig7EvtGen_H7UE_Wt_DR_inclusive
$t\bar{t}$ interference	410654, 410655	PowhegPythia8EvtGen_A14_Wt_DS_inclusive

Table 3.4: Single-top modelling cross section uncertainties.

Process	XS uncertainty
Wt	5.3%
t-channel	4.2%
s-channel	3.9%

Table 3.5: Relative acceptance uncertainties obtained for the single top Wt in each $\tau_{\text{lep}}\tau_{\text{had}}$-SLT SR and inclusively.

Region	low-m_{HH} ggF SR	high-m_{HH} ggF SR	VBF SR	Inclusive SR
PS	up: – 4.5%	up: – 4.6%	up: 1.7%	up: – 4.2%
ME	up: 6.2%	up: 3.0%	up: 20.3%	up: 5.3%
Interference	up: 6.7%	up: – 22.9%	up: – 14.0%	up: – 9.8%
PDF	up: 1.1%	up: 1.4%	up: 0.4%	up: 0.9%
ISR	up: 1.2%, do: – 1.2%	up: 4.0%, do: – 3.1%	up: 12.0%, do: – 10.5%	up: 3.2%, do: – 2.6%
FSR	up: 5.1%, do: – 4.9%	up: 2.7%, do: – 5.7%	up: 12.7%, do: – 7.8%	up: 4.3%, do: – 5.4%

Shape acceptance uncertainties are obtained based on the BDT score distribution, using the above-mentioned binning approach. The binning parameter values $t_f = 0.05$ and $t = 0.0$ are used in all cases for the Wt uncertainties. Due to statistical limitations, the value of t_s had to be adapted in order to obtain a reasonable binning in all cases. The used values are summarized in Table 3.6.

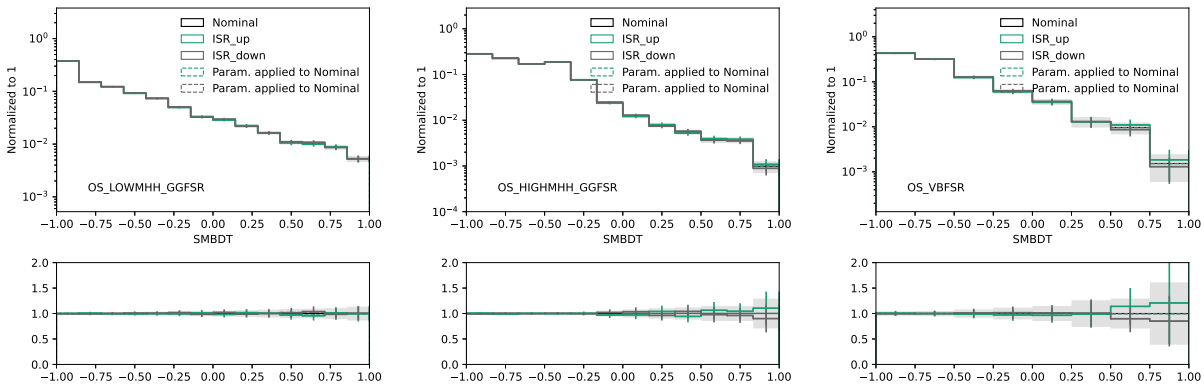
Table 3.6: Value of the shape parameterization binning parameter t_s depending on the SR and the considered variation for SLT.

Region	ISR & PS	FSR	Interference	ME
Low-m_{HH} SR	0.08	0.15	0.08	0.08
High-m_{HH} SR	0.05	0.08	0.15	0.12
VBF SR	0.08	0.15	0.08	0.08

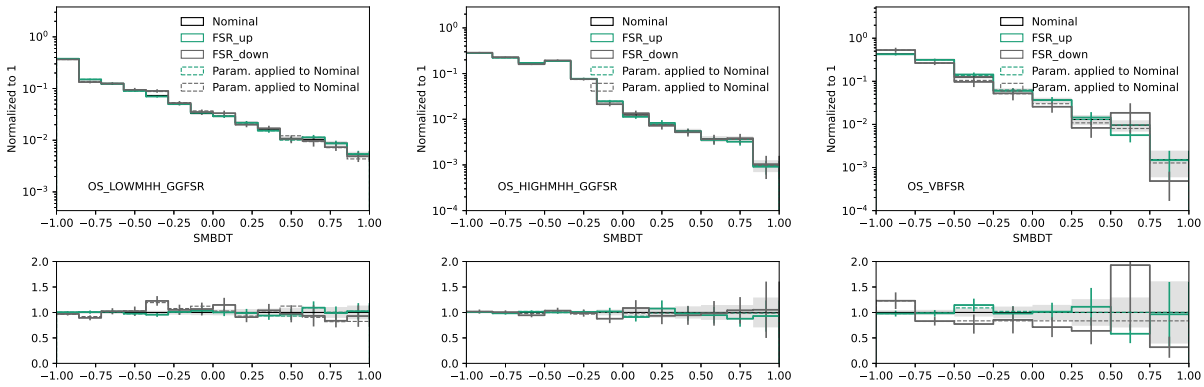
All BDT score comparisons for the nominal and varied predictions for the Wt process, including the derived parameterizations, are shown in Figures 3.17 and 3.18.

3.5.3 Modelling uncertainties on data-driven background

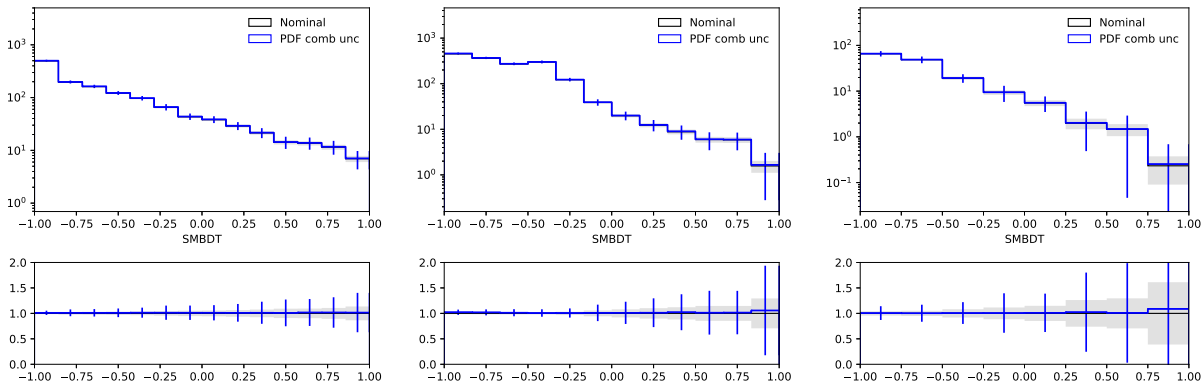
Data-driven processes contributing to the analysis selection are estimated with the techniques described in Section 3.3.1. We consider in particular background from processes



(a) low-m_{HH} ggF SR (**ISR**) (b) high-m_{HH} ggF SR (**ISR**) (c) VBF SR (**ISR**)



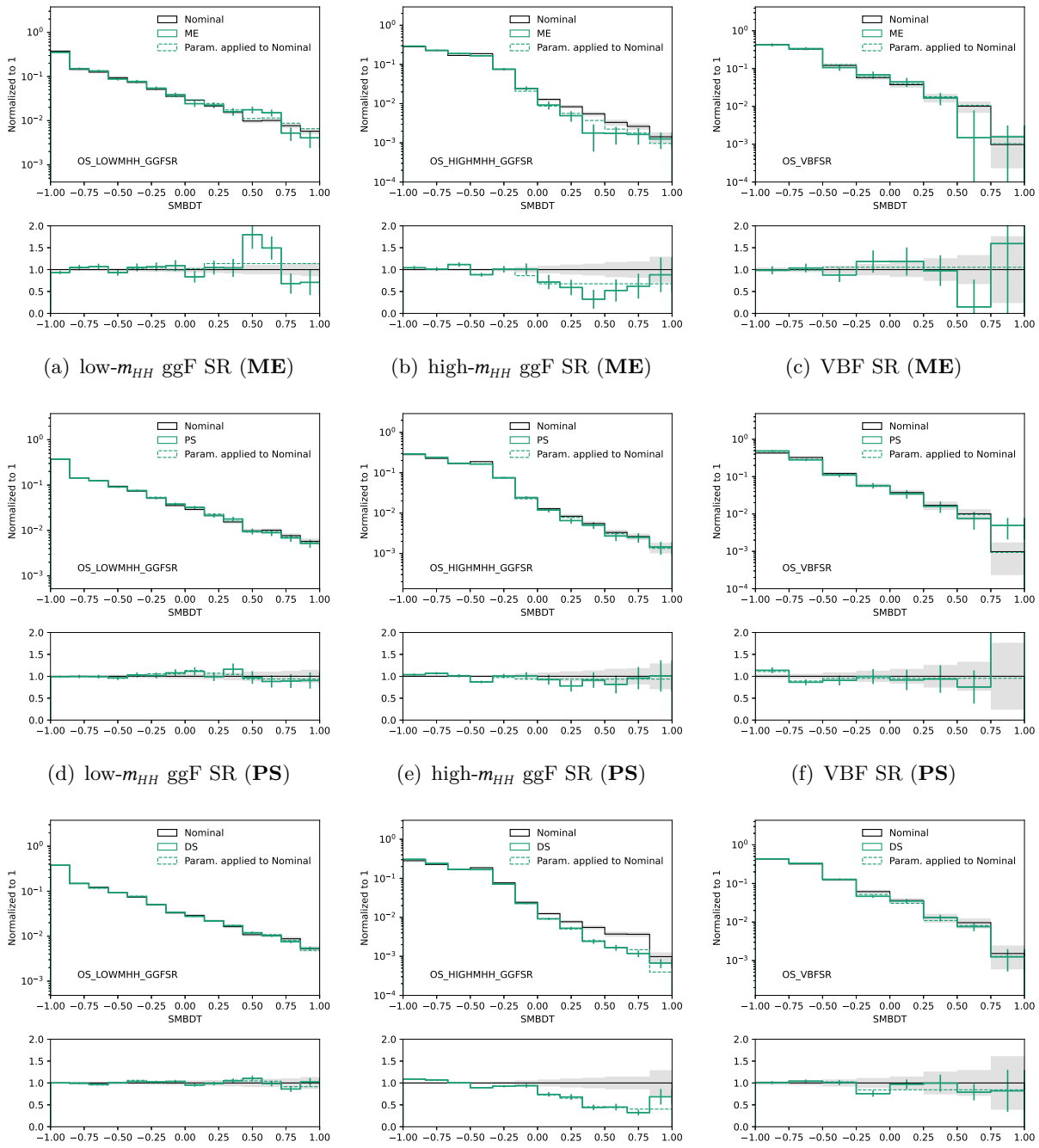
(d) low-m_{HH} ggF SR (**FSR**) (e) high-m_{HH} ggF SR (**FSR**) (f) VBF SR (**FSR**)



(g) low-m_{HH} ggF SR (**PDF**) (h) high-m_{HH} ggF SR (**PDF**) (i) VBF SR (**PDF**)

Figure 3.17: Single top Wt BDT score distribution comparisons for weight-based variations in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SRs.

containing objects mimicking the signature of hadronic τ leptons: these processes are named Fake (regardless of the background source of misidentification) in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, while in the $\tau_{\text{had}}\tau_{\text{had}}$ channel they are divided in Fake (for multi-jet processes mimicking hadronic τ signatures) and ttbarFake (for $t\bar{t}$ processes). The background estimate techniques are strongly inherited from the previous round of analysis [79]: the τ fake factors are derived as a function of fundamental properties of the taus, such as their transverse momentum and their prong-ness. Given that our new regions represent a sub-categorization of the signal



(a) low-m_{HH} ggF SR (**ME**) (b) high-m_{HH} ggF SR (**ME**) (c) VBF SR (**ME**)

(d) low-m_{HH} ggF SR (**PS**) (e) high-m_{HH} ggF SR (**PS**) (f) VBF SR (**PS**)

(g) low-m_{HH} ggF SR (**Interference**) (h) high-m_{HH} ggF SR (**Interference**) (i) VBF SR (**Interference**)

Figure 3.18: Single top Wt BDT score distribution comparisons for variations based on alternative samples in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SRs.

regions used in the previous analysis, we considered the same fake factors as in the previous round and validated that the modelling in the Fake-enriched Control Regions and Validation Regions with the new analysis categorization still provides a good description of the data distributions.

Based on this key assumption, we consider that we can rely on the same model of systematic uncertainties adopted in the previous analysis for Fake-τ_{had} backgrounds, both in the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels.

3.5.4 Modelling uncertainties on signal processes

Cross section, acceptance uncertainties (normalization and shape) for the non-resonant signal prediction closely follow the approach of the previous analysis. They are evaluated by comparing the nominal signal samples to alternative MC samples for parton shower while the effect of PDF and scale variations is assessed through internal weights in the nominal sample. The uncertainties related to the parameterization of the signal as a function of the anomalous coupling parameters are also considered.

3.6 Statistical analysis

The results of the $HH \rightarrow b\bar{b}\tau^+\tau^-$ analysis are obtained through a simultaneous binned maximum-likelihood fit to the MVA output distributions over all $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ event categories as described in Section 3.2.4 and Section 3.4, and to the m_{ll} distribution in the Z+HF CR.

3.6.1 Statistical model

The likelihood function of the analysis can be written as:

$$\mathscr{L}(\mu, \theta; \text{data}) = \prod_{c=1}^{N_{\text{cats}}} \mathscr{L}_c(\mu, \theta; \text{data}) \prod_{k \in \text{constraint NPs}} f_k(\theta_k) \tag{3.4}$$

where μ and θ are the vectors of parameters of interest (POIs) and nuisance parameters (NPs) respectively, N_{cats} is the number of analysis categories, $\mathscr{L}_c$ is the likelihood function for category c. Some classes of NPs are subject to external constraints f_k. More details about the used likelihood template can be found in [124].

The POIs are the parameters we would like to measure, e.g. the signal strength, denoted also as μ, or coupling modifiers like κ_λ and κ_{2V}. The fit model includes NPs, some of which are determined only from data and are referred to as unconstrained. Other parameters are constrained using information from auxiliary measurements in addition to the data. These parameters are referred to as constrained and quantify the effect of systematic uncertainties in the measurement.

In each category c, the likelihood is expressed in binned form as the product of Poisson distributions P, one for each analysis bin:

$$\mathscr{L}_c(\mu, \theta; \text{data}) = \prod_{i=1}^{n_{\text{bins}}} P\left(\sum_s N_{H_s}^{c,i}(\mu) + \sum_b N_{B_b}^{c,i}, n_i\right) \tag{3.5}$$

where n_i is the observed number of data events in bin i, $N_{H_s}^{c,i}$ is the yield in bin i of signal

sample s, and $N_{B_b}^{c,i}$ is the yield in bin i of background sample b. The expression appearing in the argument of the Poisson distribution, $\sum_s N_{H_s}^c(\boldsymbol{\mu}) + \sum_b N_{B_b}^c$, consequently is the predicted total event yield for bin i.

The measurement of the POIs is performed using a statistical test based on the profile likelihood ratio:

$$\Lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} \tag{3.6}$$

where μ and θ represent the POIs and the NPs respectively. In the numerator, the NPs are set to their profiled values $\hat{\hat{\theta}}$, which conditionally maximize the likelihood function of fixed values of POIs μ. In the denominator, both the POIs and NPs are set to their best-fit values $\hat{\mu}$ and $\hat{\theta}$ respectively, which maximize the likelihood unconditionally. In the asymptotic regime, in which the likelihood is approximately Gaussian, the value of $-2ln\Lambda(\mu)$ follows a χ^2 distribution with a number of degrees of freedom (d.o.f) equal to the dimension of the vector μ.

In the absence of signal, upper limits on the HH signal strength, μ_{HH}, and on the di-Higgs production cross-section at 95% CL are set following the CL_s prescription [125] and using the asymptotic formula. The considered test statistic $\tilde{q}_\mu$ is defined as:

$$\tilde{q}_\mu = \begin{cases} -2\ln\frac{\mathscr{L}(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathscr{L}(0,\hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} < 0, \\ -2\ln\frac{\mathscr{L}(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathscr{L}(\hat{\mu},\hat{\boldsymbol{\theta}})} & 0 \le \hat{\mu} \le \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{3.7}$$

3.6.2 Fit setup

The binned profile likelihood fit is performed simultaneously on the MVA scores in all analysis regions, together with the m_{ll} distribution in the Z+HF CR, as summarized in Table 3.7.

Table 3.7: Regions entering the fit and fitted observable in each analysis region and channel.

	low-m_{HH} SR	high-m_{HH} SR	VBF SR	Inclusive
$\tau_{\text{had}}\tau_{\text{had}}$	BDT score	BDT score	BDT score	–
$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	BDT score	BDT score	BDT score	–
$\tau_{\text{lep}}\tau_{\text{had}}$ LTT	BDT score	BDT score	BDT score	–
Z+HF CR	–	–	–	m_{ll}

The POI for the SM fits is the HH signal strength, μ_{HH}, which is relative to the ggF+VBF input signal cross-section of $31.05 + 1.726$ fb times the BR. Fits are also performed with two POIs, μ_{ggF} and μ_{VBF}, for the ggF and VBF HH signal strength respectively, each of them having its own normalization factor. The normalization of the $t\bar{t}$ and Z+HF jets backgrounds

are freely floating in the fit and are determined from data. Relative acceptance uncertainties between the SRs and CR are applied to these normalizations in the SRs included in the fit.

For the κ_λ and κ_{2V} likelihood scans, the signal MVA score distributions are parameterized in terms of κ_λ and κ_{2V} respectively, by producing 3 ggF and 6 VBF signal templates, as described in Section 1.2.4. The workspaces are modified to correlate the signal samples through the new POIs κ_λ and κ_{2V}. The signals are combined to obtain signal distributions for arbitrary κ_λ and κ_{2V} values respectively using the linear combination method introduced in Section 1.2.4. The single-Higgs production cross-sections and BRs are also varied with κ_λ in the fit.

Templates corresponding to similar physical processes are merged into a single template in the fit. All sources of systematic uncertainties are considered as NPs in the profile likelihood fit. The effect of each NP is split into normalization and shape components, corresponding to the normalization and shape uncertainties respectively as discussed in the previous section. The shape uncertainties are included in the form of alternative histograms for the fit templates. Normalization uncertainties are implemented as either flat (for "floating" templates) or Gaussian priors.

Below are listed the correlations of NPs across all fit regions in a general perspective:

- All experimental uncertainties are correlated across all fit regions
- Cross-section and acceptance uncertainties on the MC estimated backgrounds and on signal are also correlated
- Floating $t\bar{t}$ and Z+HF normalizations are correlated across all fit regions
- Relative acceptance uncertainties on the normalization of $t\bar{t}$ and Z+HF, which are determined from data in the fit, are correlated with the shape variations from the same source of uncertainty between the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels. The main exception from the rule is the shape uncertainties due to generator comparison for Z+hf in the CR which is kept uncorrelated to extrapolation and shape uncertainties in the SR in order to avoid propagating the observed constrain likely due to a fluctuation in the alternative sample distribution.
- Uncertainties on the data-driven backgrounds are not correlated

The NPs undergo a series of treatments before being added to the fit model. These include symmetrization, smoothing and pruning. Symmetrization:

- One-sided experimental and modelling systematics are symmetrized
- Experimental systematics resulting in same-sided variations are also symmetrized, using the average of up and down variations. This is done to fix under-constraints from same-sided variations

The systematic variations can also be smoothed in order to avoid the instabilities in the likelihood minimization which arise from the statistical fluctuations in the templates.

Smoothing is applied on 4-vector-based CP variations, but not on weight CP variations. Several algorithms can be used for smoothing. The algorithm that is used by default is based on an iterative re-binning of the systematic variations until a certain number of local extrema is attained. By default, two extrema are used, corresponding to systematic variations which are monotonic in the fitted variable.

For the shape variations arising from theory uncertainties, this smoothing algorithm is not used. This is because they have already been derived using a dedicated smoothing algorithm that ensures e.g. that a minimum number of background event is present in each bin to derive the variation. The details of this algorithm are explained in Section 3.5.2.

Finally, the pruning is applied to all systematics for the purpose of speeding up the fitting process and also reducing instabilities that arise from small systematics that are only introducing noise and have a negligible impact on the analysis sensitivity. Pruning is applied separately on the normalization and the shape effects of a systematic uncertainty. The pruning criteria applied are as follows:

- the normalization effect of a systematic uncertainty is pruned away if the difference between the nominal and systematic yield is less than 0.5%
- the shape effect of a systematic uncertainty is pruned away if the variation between the nominal and the systematic templates is less than 0.5% in all bins
- the shape systematics having up and down variations on the same side are pruned away based on a χ^2 test; if the χ^2 value between up and down variation is smaller than the χ^2 between nominal and up/down variation (only applied to the smoothed systematics)

Additionally, MC statistical uncertainties are also included as NPs with Poissonian priors.

3.6.3 Binning

The MVA output score distributions are first built with a very fine binning and are then re-binned to build the fit templates. The goal is to minimize the number of bins, while also maximizing the retained expected sensitivity, and ensuring the stability of the fit and validity of the asymptotic approximation. The same re-binning algorithm (so-called Trafo60) is adopted by both $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels and is described below.

All BDT score histograms are built with 2090 bins following a non-uniform binning. The scores are filled from -1 till 0.990 with a 10^{-3} bin width (1990 bins), while an even finer width of 10^{-4} is used from 0.990 to 1 (100 bins), corresponding to the most signal-like scores. The re-binning procedure starts from these finely binned histograms, and bins are then iteratively merged, beginning from the most signal-like MVA bins until certain criteria are fulfilled.

The following general function can be used to transform the BDT output histograms:

$$Z(I[k,l]) = Z(z_s, n_s(I[k,l]), N_s, z_b, n_b(I[k,l]), N_b)\,, \tag{3.8}$$

where

- $I[k,l]$ is an interval of histograms, containing the bins between bin k and bin l
- N_s is the total number of signal events in the histogram
- N_s is the total number of background events in the histogram
- $n_s(I[k,l])$ is the total number of signal events in the interval $I[k,l]$
- $n_b(I[k,l])$ is the total number of background events in the interval $I[k,l]$
- z_s and z_b are parameters used to tune the algorithm.

In our analysis, the above function takes the following form:

$$Z = z_s \frac{n_s}{N_s} + z_b \frac{n_b}{N_b}\,, \tag{3.9}$$

where $(z_s, z_b) = (10, 5)$ is chosen for the BDT in all the analysis regions. The value for z_s is higher than z_b to achieve finer binning in the high BDT output score regime which has a very high signal-to-background ratio. Starting from the last bin on the right of the original histogram (high-BDT score), we increase the range of the interval $I[k, \text{last}]$ by adding one bin after the other, from right to left. The value of Z is calculated at each step. Once $Z(I[k_0, \text{last}]) > 1$, all bins in the interval $I[k_0, \text{last}]$ are merged into a single bin, where "last" here refers to the rightmost bin of the original hist in the first iteration. After each iteration, the last bin is the rightmost one excluding the newly formed bins from previous iterations.

Furthermore two additional requirements are applied after each iteration and the above steps are repeated until these are fulfilled:

- the MC statistical uncertainty on the sum of backgrounds of each bin to be lower than 20%.
- the number of expected background events in each bin is required to be greater than 3.[4]

The requirement of at least three expected background event per bin is more aggressive than the suggestion in the HDBS group to have at least five expected background events per bin in order to trust the asymptotic approximation in the inference procedure. Therefore, dedicated checks have been performed in the $\tau_{\text{had}}\tau_{\text{had}}$ channel to compare limits obtained with toys with the values obtained when using the asymptotic approximation. These checks were done at an earlier stage of the analysis and only included floating normalizations and MC statistical uncertainties in the fit. In all cases, the difference between the asymptotic

[4]With respect to previous studies, the requirement has been increased from 1 to 3 to ensure that the major backgrounds are sufficiently populated in the most discriminant BDT bins.

approximation and the limit obtained from toys was smaller than 1.5% and similar to the difference obtained when using the previous analysis workspace. The team plans to redo these studies once the systematics model is finalized.

3.7 Results

The simultaneous fit is performed including all the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ SRs, and the Z+HF CR as shown in Table 3.7. The CL_s method [125] is employed to determine the upper limits of signal strength. Upper limits are calculated for the overall μ_{HH} and for the individual signal strength parameters μ_{ggF} and μ_{VBF}. The constraints on the coupling modifiers are established by utilizing a profile likelihood ratio $\Lambda(\alpha, \theta)$ as the test statistic, calculated from the likelihood function in the asymptotic approximation, with the coupling modifiers κ_λ and κ_{2V} being the parameter of interest (POI) α. The expected results are achieved by utilizing Asimov datasets produced from the likelihood function after assigning the maximum likelihood estimate to all nuisance parameters in the fit to data, while also fixing the values of the POIs to correspond to the hypothesis being tested. Figure 3.19 displays the distribution of BDT scores across the nine orthogonal categories after the maximum likelihood fit to the data.

The combined fit to data gives the total HH production signal strength of $\hat{\mu}_{HH} = 2.2 \pm 1.7$ and the $t\bar{t}$ and Z+HF normalisation factors of respectively 0.96 ± 0.03 and 1.34 ± 0.08. The observed 95% confidence level upper limit on μ_{HH} is 5.9 while the expected one is 3.3 under the background-only assumption. In the simultaneous fit of μ_{ggF} and μ_{VBF}, the observed (expected) 95% confidence level upper limits are respectively 5.8 (3.4) and 91 (73) for each production mode. In the one-dimensional fit of μ_{ggF} where μ_{VBF} is set to its SM prediction, the the observed (expected) 95% confidence level upper limit is 5.9 (3.4). Likewise, In the one-dimensional fit of μ_{VBF} where μ_{ggF} is set to its SM prediction, the observed (expected) 95% confidence level upper limit is 93 (72). It's worth mentioning that, the expected upper limits on the separate production mode signal strengths are determined under the background-only assumption.

The upper limits on signal strength are shown in 3.8 and 3.20. The results for the individual SRs are acquired through the combined likelihood fit of the BDT score distributions in the categories of a single SR, along with the $m_{\ell\ell}$ distribution from the dedicated CR. The observed limit on μ_{HH} from the combined fit is looser than the expected one because of an excess in the high-m_{HH} category, in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR. From the individual fit of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR, this excess corresponds to a local significance of 2.3 σ with respect to the SM hypothesis ($\mu_{HH} = 1$).

Figure 3.21 shows the expected and observed scans of the negative log-likelihoods (NLL) as a function of κ_λ and κ_{2V}, assuming that all the other coupling modifiers are fixed to SM

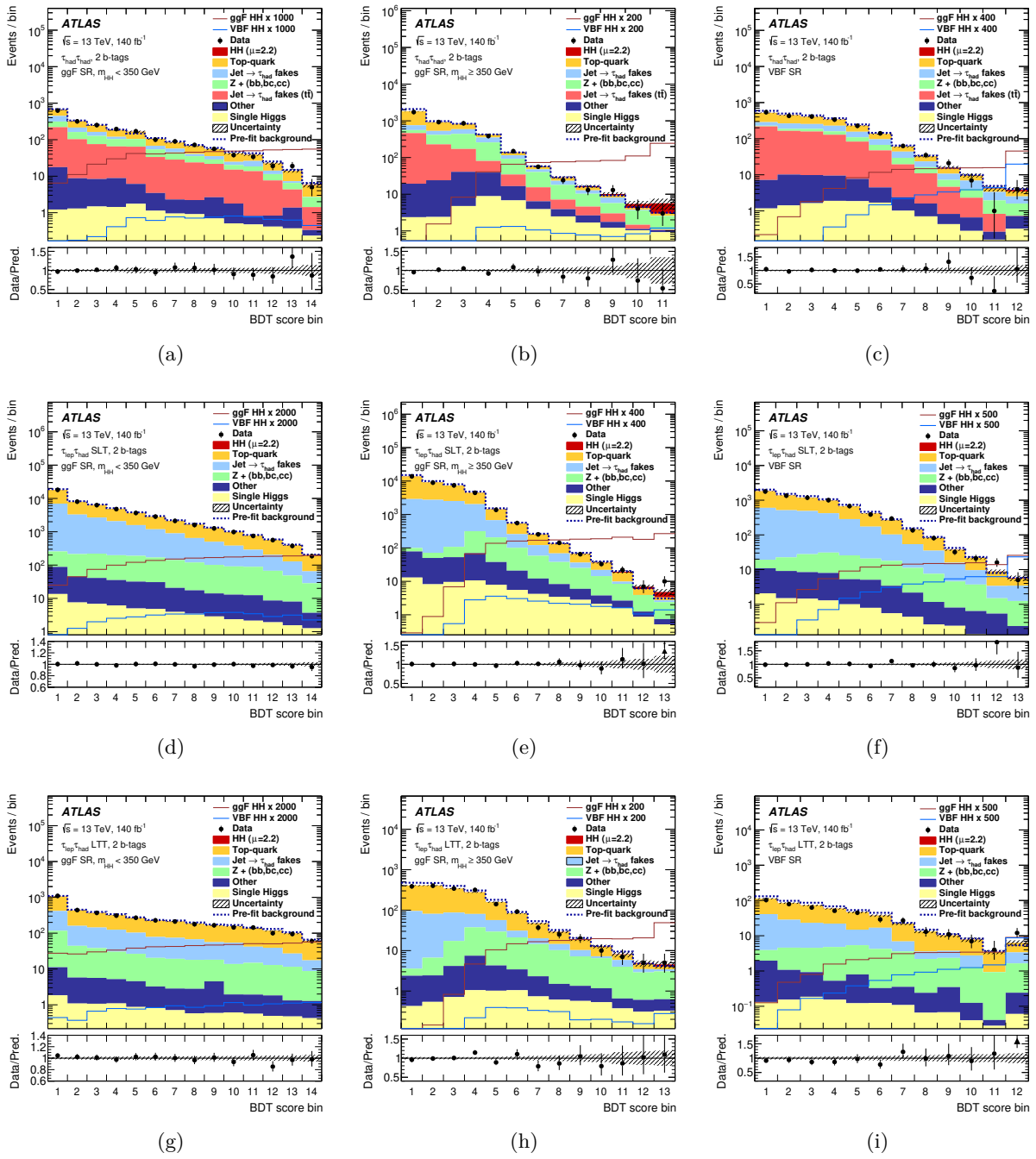


Figure 3.19: Post-fit BDT score distributions in the low-m_{HH} (left), high-m_{HH} (middle column) and VBF (right) categories of the $\tau_{\text{had}}\tau_{\text{had}}$ (top), $\tau_{\text{lep}}\tau_{\text{had}}$ SLT (middle row) and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT (bottom) SRs. The contribution of the SM HH signal is adjusted based on the fitted signal strength μ_{HH} obtained from the combined likelihood fit multiplied by the SM expectation. The ggF and VBF HH signal distributions are superimposed and adjusted according to the scaling factor specified in the legend multiplied by the SM expectation. The dashed histograms represent the total pre-fit background. The lower panels display the ratio of data to the total post-fit sum of signal and background, with the hatched bands indicating the statistical and systematic uncertainties of this estimate. The BDT score distributions are presented using the same binning as in the likelihood fit. Each bin is visualized with a consistent width for better visualization, and the x-axis denotes the bin number.

Table 3.8: The observed and expected 95% confidence level upper limits on μ_{HH}, μ_{ggF} and μ_{VBF} were determined by performing the likelihood fits of the individual SRs, as well as the combined likelihood fits. The limits for μ_{ggF} and μ_{VBF} were derived from both the simultaneous fit of both signal strengths and separate fits for each production mode, assuming the other to be SM-like. The uncertainties associated with the combined expected upper limits relate to the 1σ uncertainty band.

		μ_{HH}	μ_{ggF}	μ_{VBF}	μ_{ggF} (μ_{VBF}=1)	μ_{VBF} (μ_{ggF}=1)
$\tau_{\text{had}}\tau_{\text{had}}$	observed	3.4	3.6	87	3.5	80
	expected	3.8	3.9	102	3.9	99
$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	observed	17	17	136	17	158
	expected	7.2	7.4	129	7.4	127
$\tau_{\text{lep}}\tau_{\text{had}}$ LTT	observed	23	18	765	22	733
	expected	20	21	359	20	350
Combined	observed	5.9	5.8	91	5.9	93
	expected	$3.3^{+1.7}_{-0.9}$	$3.4^{+1.8}_{-1.0}$	73^{+32}_{-21}	$3.4^{+1.8}_{-0.9}$	72^{+32}_{-20}

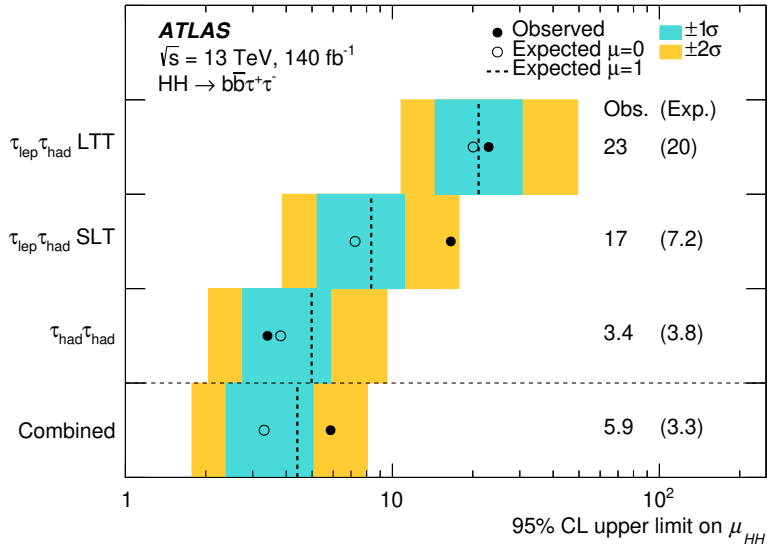
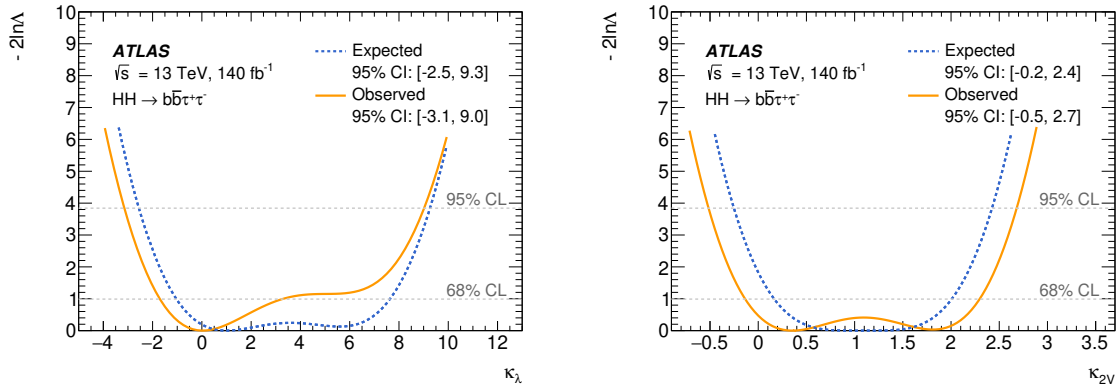


Figure 3.20: Summary of observed (filled circles) and expected (open circles) 95% confidence level upper limits on μ_{HH} from the fit of each individual channel and the combined fit in the background-only ($\mu_{HH} = 0$) hypothesis. The dashed lines indicate the expected 95% confidence level upper limits on μ_{HH} in the SM hypothesis ($\mu_{HH} = 1$). The cyan and yellow bands indicate the $\pm1\sigma$ and $\pm2\sigma$ variations on the expected limit with respect to the background-only hypothesis due to statistical and systematic uncertainties, respectively.

values. The observed (expected) 95% confidence intervals of κ_λ and κ_{2V} are respectively $-3.1 < \kappa_\lambda < 9.0$ ($-2.5 < \kappa_\lambda < 9.3$) and $-0.5 < \kappa_{2V} < 2.7$ ($-0.2 < \kappa_{2V} < 2.4$). Additionally, κ_λ and κ_{2V} are constrained simultaneously while assuming other coupling modifiers are set to one. Figure 3.22 shows the observed and expected two-dimensional 68% and 95% contours.



(a) (b)

Figure 3.21: NLL scans for κ_λ (a) and κ_{2V} (b) acquired from the fits to data (orange) and an Asimov dataset (dashed blue) built from the SM assumption. For both cases, all coupling modifiers except for the scanned parameters are set to the SM predictions.

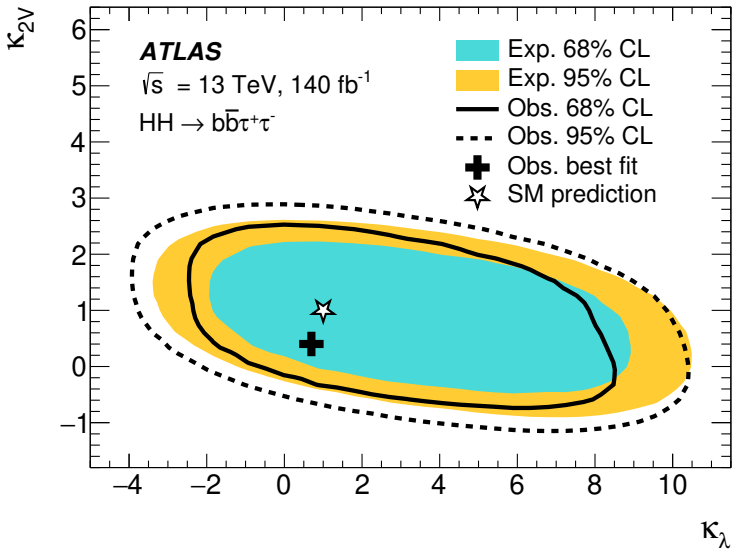


Figure 3.22: The 68% (solid line) and 95% (dashed line) confidence level likelihood contours in the (κ_λ, κ_{2V}) plane, while setting the other coupling modifiers to one. The cyan and yellow shaded regions display the expected contours. The star represents the SM prediction, while the black cross represents the best-fit value.

As mentioned in the previous round of analysis [79], the sensitivity of the analysis is primarily limited by the statistical uncertainty of the data. The main systematic uncertainty affecting the measurement of μ_{HH} is the uncertainty in the ggF HH production cross-section due to variations in the QCD scales and the top-quark mass scheme. Following this, the next significant sources of uncertainty include the statistical precision of the background MC samples and the uncertainty associated with the interference between the Wt and $t\bar{t}$ processes. The collective impact of all systematic uncertainties results in a 23% increase in the expected upper limits on the signal strength μ_{HH} and the 9% and 2% widening of the expected 95% confidence interval respectively for κ_λ and κ_{2V}.

Table 3.9: Comparison between expected and observed quantities for the current and previous analysis round in the combined analysis.

analysis	95% CL upper limit on μ_{HH}	95% CI for κ_λ	95% CI for κ_{2V}
current analysis (Obs.)	5.9	[-3.1, 9.0]	[-0.5, 2.7]
previous analysis (Obs.)	4.7	[-2.7, 9.5]	[-0.6, 2.7]
diff.	+25%	-1%	-2%
current analysis (Exp.)	3.3	[-2.5, 9.3]	[-0.2, 2.4]
previous analysis (Exp.)	3.9	[-3.07, 10.20]	[-0.54, 2.67]
diff.	-15%	-11%	-19%

Due to the usage of a consistent statistical procedure, a comparison can be made between these results and the ones in the previous round of analysis [79]. Table 3.9 shows a comparison between expected and observed quantities for the current and previous analysis round in the combined analysis. As can be seen, the strategy adpoted in this analysis results in a 25% (15%) increase (reduction) in the observed (expected) upper limit on the signal strength μ_{HH}; The observed (expected) confidence intervals for κ_λ and κ_{2V} are respectively reduced by 1% (11%) and 2% (19%).

Chapter 4

The ATLAS Phase-2 Upgrade

This chapter starts with a general introduction of the physics motivation about proceeding to the HL-LHC, followed by an overview of the HL-LHC project. Then, the ATLAS Phase-2 upgrade is briefly described, focusing on how each sub-detector is upgraded based on the new requirements. Finally, the new detector HGTD is introduced from the aspects of hybrid module, peripheral electronics, DAQ, DCS and luminosity.

4.1 Physics motivation

The Higgs boson self-coupling is one of the most important Higgs properties in the Standard Model. It can help us to have a better understanding of the structure of the Higgs potential, which is the key to the Higgs mechanism. The Higgs mechanism then opens the door to masses for other Standard Model particles. Further, the electroweak spontaneous symmetry breaking is introduced by the Higgs mechanism. Both ATLAS and CMS experiments have already made a big effort on the measurement of the Higgs boson self-coupling through the processes of single- and double-Higgs production. Due to the limited amount of dataset recorded so far, only a rather loose constraint on the Higgs boson self-coupling can be set, which is pretty far away from having a direct measurement of it. For example, the result in Chapter 3 shows the latest constraint on the Higgs boson self-coupling with the ATLAS Run-2 dataset through one of the most sensitive final states of the double-Higgs decay, $b\bar{b}\tau^+\tau^-$. As can be seen, although the expected constraint with 95% confidence level ($-2.5 < \kappa_\lambda < 9.2$) is improved by about 11% compared with previous analysis, it is still far from enough. There are even some efforts on combining the three most sensitive final states of double-Higgs decay ($b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$ and $b\bar{b}\gamma\gamma$) [7], and also the efforts on combining not only the three most sensitive final states but also some important final states of single-Higgs decay [126]. Those combination results indeed get improved, but they are still at the level where the Higgs boson self-coupling can only be constrained within a large range. Even with the Run-3 data considered, the results on Higgs boson self-coupling wouldn't change significantly and we would still be in the same situation.

At the current stage of the LHC, the lack of dataset is the main limiting factor that stops us from performing a more decent measurement on the Higgs boson self-coupling. If more data is collected, it will certainly improve the measurement. In fact, the dataset expected at the end of LHC Phase-1 is just a small fraction ($< 10\%$) of the dataset planned for the whole LHC lifetime, and another more than 90% of dataset will be collected during the HL-LHC (LHC Phase-2). The HL-LHC will produce approximately 170 million Higgs bosons and 120,000 Higgs-boson pairs on each of the ATLAS and CMS experiments over a period of about 10 years. With the tenfold Higgs boson events of HL-LHC, the Higgs boson self-coupling will be significantly constrained. A HL-LHC projection study [127] shows that, the 95% confidence level constraint can be narrowed down to $-0.1 < \kappa_\lambda < 2.7 \cup 5.5 < \kappa_\lambda < 6.9$ by combining the three most sensitive final states of double-Higgs decay ($b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$ and $b\bar{b}\gamma\gamma$) at ATLAS experiment. There is also another HL-LHC projection study [128], showing that the Higgs boson self-coupling strength will be further constrained to $-0.1 < \kappa_\lambda < 2.3$ with 95% confidence level by combining the major final states of double-Higgs decay from both the ATLAS and CMS experiments, and a significance of 4 times standard deviation can be achieved with all systematics uncertainties included, see Figure 4.1 for more details. As can be seen, the significantly improved measurement on the Higgs boson self-coupling can be one of the strong motivations of HL-LHC.

In addition to the measurement on the Higgs boson self-coupling, there is another measurement that must be mentioned, which also benefits to a great extend from the increased dataset of HL-LHC. It is the precision measurement of Higgs boson properties, such as cross sections, branching ratios, coupling parameters and mass. According to the projection study [129] from ATLAS Run-2 dataset, $2.4 - 7.7\%$ precision on the measurement of Higgs boson production cross sections can be reached and $1.8 - 4.3\%$ precision on the couplings to W, Z and third-generation fermions can also be reached, see Figure 4.2 for more details.

With the dataset accumulated up to 3000 fb^{-1}, each analysis group might be able to push boundary to a higher limit where the SM can be further consolidated, a new theory like the SUperSYmmetry (SUSY) [130] can be confirmed, particle properties are measured with high precision, or there might be some breakthroughs on some unanswered questions, such as dark matter, naturalness [131] and baryogenesis. Therefore, there are more other physics programs that draw great attention from physicists and can be performed at the HL-LHC, which can be generally summarized into different topics, such as Standard Model measurements, studies of the properties of the Higgs boson, searches for phenomena beyond the Standard Model, flavor physics of heavy quarks and leptons, and studies of QCD matter at high density and temperature. With all those physics motivations, the HL-LHC is already on the spotlight and scientists at CERN have been already working hard on the schedule of making it happen.

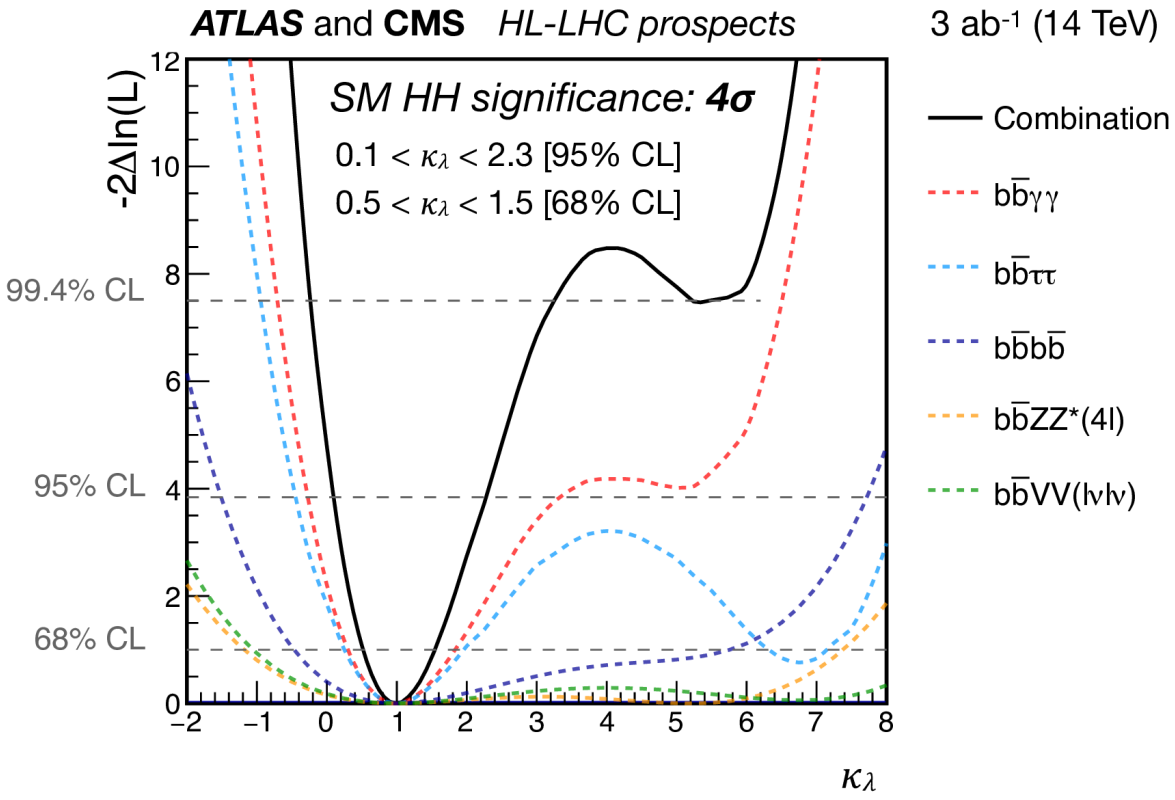


Figure 4.1: Negative-log-likelihood scan as a function of κ_λ, calculated by performing a conditional signal+background fit to the background and SM signal. The colored dashed lines correspond to the combined ATLAS and CMS results by channel, and the black line their combination. [128].

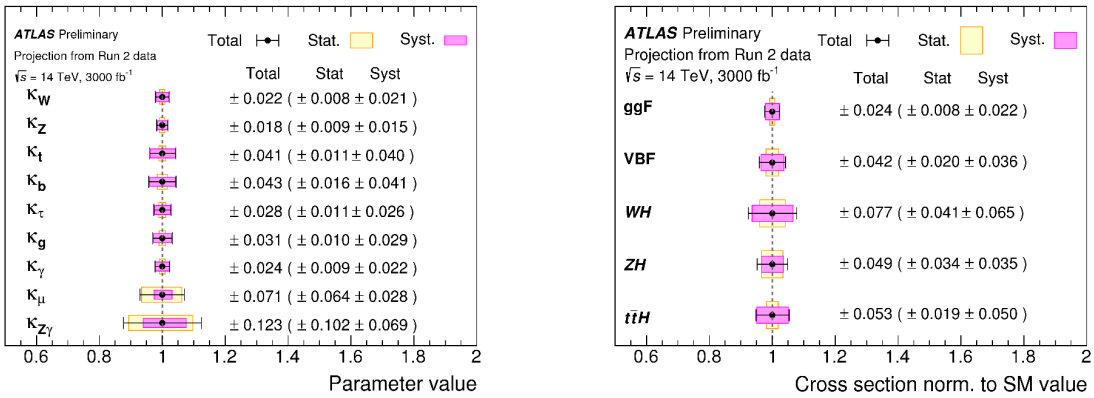


Figure 4.2: Expected result (a) for the measurement of each Higgs boson coupling modifier per particle type, and expected result (b) for the measured cross sections for the ggF, VBF, WH, ZH and tt̄H production modes normalised to their SM predictions. The systematic uncertainties are assumed to be half of the current Run-2 uncertainties [129].

4.2 HL-LHC

In order to fully exploit the physics potential of the LHC, CERN established the HL-LHC project at the end of 2010. With major technical upgrades, the HL-LHC aims for the extension of operability and an increase of the collision rate, which means the ability of producing more events of interest for the physics analyzers. Figure 4.3 shows the schedule of the LHC and HL-LHC. The HL-LHC is expected to deliver beam collisions at the start of 2029 with the peak levelled instantaneous luminosity of $5 \times 10^{34} cm^{-2} s^{-1}$, five times the LHC nominal luminosity. If operated in this level of luminosity, it can produce an integrated luminosity of 250 fb^{-1} per year, and realize the goal of 3000 fb^{-1} over the course of twelve years or so after the upgrade. By further using the engineering margins reserved in the equipment design for resisting the instantaneous heat deposition and the integrated radiation dose [132], the HL-LHC is able to reach an ultimate performance of instantaneous luminosity $7.5 \times 10^{34} cm^{-2} s^{-1}$. If operated in this level of luminosity, it is able to produce an integrated luminosity of 300 to 400 fb^{-1} per year, thus yielding about 4000 fb^{-1} dataset. The more intense and concentrated beam makes the pile-up (the number of collisions that happen in one bunch crossing) reach to a higher level. The average pile-up will be around 200, and it is around 50 for the Run-3 data taking.

Due to the higher requirement on the beam of the accelerator, some equipments of the LHC have to be upgraded. The main updates and layout modifications for HL-LHC will be carried out in the 600-meter long insertion regions of Point 1 (P1, near ATLAS experiment) and Point 5 (P5, near CMS experiment). Other parts that require modifications sum up to about 2.2km. In the insertion regions of P1 and P5, almost a complete renewal will be performed by using the Q4 quadrupole and new low-β inner triplet quadrupoles, also there will be an installation of one large 1.9K refrigerator, installation of new collimators and insertion of crab cavities in the matching sections. A new equipment for beam monitoring, such as beam position monitoring and beam gas vertex profile monitoring, will also be installed. Additionally, the major civil engineering works of both underground and surface will happen at P1 and P5, and some dedicated underground caverns and galleries with the total length of one kilometer will be made accessible even during the LHC operation with beam.

Several new key technologies [133] have been developed to support the upgrade of the HL-LHC equipments. A novel superconducting compound based on niobium and tin will be used to make more powerful quadrupole magnets, which can achieve way higher magnetic fields than the niobium-titanium alloy used by the current LHC magnets (12 Tesla as opposed to 8 Tesla). The beam optics has gone through a significant development such that the instantaneous luminosity can be levelled at a constant value throughout the lifespan of the beam. The crab cavities is an innovative superconducting equipment that will be installed

to give the head and tail of the particle bunches a transverse deflection while leaving the center of the bunch undeflected, and this will enlarge the overlap of two bunches and thus increase the collision rate. New collimators are being developed based on a material that brings less electromagnetic interference to the beam, and it also has the purpose of absorbing stray particles from the beam trajectory and thus provide protection for the machine. An innovative superconducting transmission line made of magnesium diboride is able to carry currents of record intensities (up to 100000 Amperes) and work at a much higher temperature, and it will be used to connect the power converters to the new accelerator magnets. In addition, new crystal collimators has been developed to boost the cleaning efficiency during ion beam operation.

In addition to the upgrades on the LHC itself, the detectors also need to be upgraded to cope with the significantly-increased pile-up. For the CMS detector, there are several parts that require upgrades [134]. The whole trigger and data acquisition (DAQ) system will be replaced to achieve a L1 trigger rate of 750 kHz and a HLT trigger rate of 7.5 kHz. In terms of the muon detector, the electronics for the drift tube and the cathode strip chamber systems will be upgraded while new detectors will also be installed in the very forward region. In terms of the calorimeter, the barrel calorimeter detectors will be remained unchanged but most of its electronics will be replaced, and the endcap calorimeter will be completely replaced by a new one called high granularity calorimeter (HGCAL). Another new detector is the MIP timing detector (MTD) precisely measuring the time of minimum ionizing particles (MIP) to improve pile-up rejection. The ATLAS Phase-2 upgrade is described in the following section.

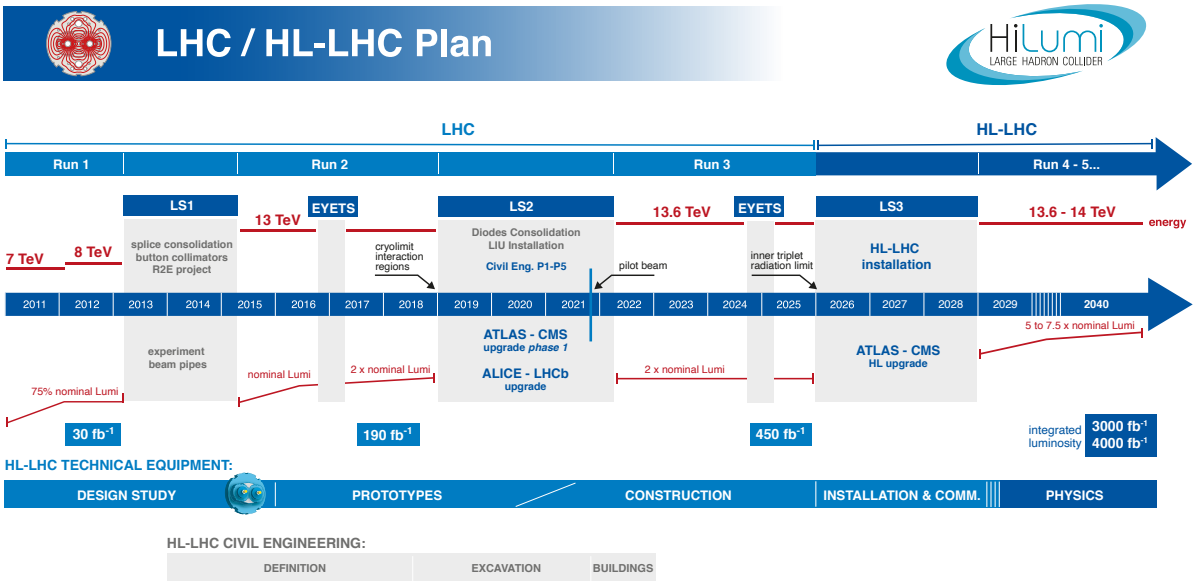


Figure 4.3: LHC/HL-LHC Plan [135].

4.3 Overview of the ATLAS Phase-2 upgrade

In order to deal with the challenges posed by the HL-LHC, the ATLAS detector has to go through a corresponding upgrade, called ATLAS Phase-2 upgrade [136]. There are two main aspects that need to be considered. One is the flood of incoming data delivered by the more powerful HL-LHC, and the ATLAS should be able to recognize and record the events of interest from the messy collisions. Another one is the increased irradiation level, which has to be taken into account throughout the development of the detector systems. That doesn't mean the ATLAS detector should be completely renewed. For example, those sub-detectors that are far away from the collision point and will not suffering from the irradiation damage during the HL-LHC operation, such as the LAr calorimeter, tile calorimeter and the muon spectrometer, will remain unchanged, but their electronics have to be upgraded more or less. Other upgrades include installations of the new inner tracking detector (ITk), high granularity timing detector (HGTD) and new muon chambers. In addition, the trigger and data acquisition (DAQ) systems will be fully upgraded.

There will be more events that need to be picked up by the detector within a certain amount of time than that of Phase-1 data taking, so the trigger system has to be updated accordingly to be able to tell the detector to record those more events [137]. The hardware-based Level-0 (L0) trigger collects the data from the calorimeters and muon spectrometer at the rate of $40\ MHz$, and make the first decision by outputting a trigger signal at the rate $1\ MHz$ (as opposed to $100\ kHz$ in the LHC period) with the latency of $10\ \mu s$, then the trigger signal will be fed back to sub-detectors. With the use of the new global trigger component, the granularity of the detector can be fully exploited. Another important layer of the trigger system is the high level trigger (HLT), which can be taken as a software-based event filter. It gets the $1\ MHz$ data flow from the sub-detectors that are just got triggered by the L0 output signal, and make a further decision by outputting another trigger signal at the rate $10\ kHz$ (as opposed to $3\ kHz$ in the LHC period). It's already decided that commercial hardware will be used for implementing the HLT. Obviously, the high trigger directly leads to a huge data throughput, which requires the upgrade of the ATLAS DAQ system. The main update of the system is its backend electronics, which will be based on a custom PCIe FPGA card (FELIX) instead of the previously used VME-based readout board.

The on-detector and off-detector electronics of both the LAr and tile calorimeters will be upgraded, and they will provide full digital input to the ATLAS trigger system and have continuous readout at the rate of 40 MHz without on-detector buffering. Some highlights of them are mentioned as follows. In terms of the LAr calorimeter [138], a new high precision frontend electronics are being developed for the on-detector electronics to achieve the dynamic range of 16 bits and a linearity of better than 0.1%, which actually benefit from the newly-developed ASICs for this purpose. Also a dedicated advanced telecommunications

computing architecture (ATCA) board are being developed for the off-detector electronics to do the waveform feature extraction. In terms of the tile calorimeter [139], those photomultiplier tubes (PMTs) that are mostly exposed to irradiation will be replaced, and they are accounting for 10% of all PMTs. In order to improve the response stability, the passive PMT high-voltage dividers will be replaced by the active ones. In addition, a new modular mechanical design for the electronics system is performed to achieve better accessibility and maintenance, and increase redundancy.

There will be a complete replacement of the current inner detector by a new all-silicon inner tracker system (i.e. ITk) [140] with a large angular coverage of $|\eta| < 4$ (2.5 for inner detector). The ITk is generally consisted of three compartments, the pixel inner system, the outer barrel and the outer endcap. The pixel inner system will be replaced at half lifetime due to the performance degradation caused by high irradiation dose. From inside to outside, the outer barrel can be further divided into pixel part and strip part, and similarly the outer endcap can also be divided into pixel part and strip part. For all the pixel parts, they cover up to $|\eta| < 4$ with 5 barrel layers and several EC disks. There are up to 5 billions active pixels for responding the particle hits, and the pixel sizes are $25 \times 100\ \mu m^2$ for the innermost barrel and $50 \times 50\ \mu m^2$ for the pixels everywhere else. It is interesting to mention that all those pixels sum up to $13\ m^2$ of silicon, which is almost 10 times larger than the current one. The smaller size of the pixels improves the resolutions of the tracking. For all the strip parts, they cover up to $|\eta| < 2.7$ with 4 barrel layers and 6 EC disks. All the active strips sum up to $168\ m^2$ of silicon, which is almost 3 times larger than the current one. The ITk has smaller interference on the particles that pass through it due to the reduced material budget. The optimized layout of the pixels and strips can make sure that a track can leave at least 9 silicon hits, which can improve the tracking and vertexing efficiency. Located in the place nearest to the collision point, the detector is designed to be extremely irradiation-tolerant, for example, the pixel inner system is required to withstand the $1\ MeV$ neutron equivalent flux of up to $1 \times 10^{16}\ n_{eq} cm^{-2}$.

The upgraded parts of the muon spectrometer are as follows [141]. The readout and trigger electronics will be renewed such that it can provide continuous readout at the rate of $40\ MHz$ and the MDT will also be able to join the L0 trigger inputs together with its other trigger logics. Additional barrel layers of sMDT, RPC and TGC will be installed in the inner part of the muon spectrometer, and by working together, they can significantly improve the coverage, trigger uniformity, momentum resolution, etc. The upgrade of the HGTD is presented in more details in Section 4.4.

4.4 HGTD

4.4.1 Overview

The large increase of pile-up is one of the biggest experimental challenges that the Phase-2 upgrades of all the experiments have to cope with. In order to mitigate the adverse effects of pile-up, a high granularity timing detector (HGTD) is therefore proposed for the ATLAS Phase-2 upgrade [142]. Based on the low gain avalanche detector (LGAD) technology, it can precisely measure the timing information of the track hits and finally distinguish collisions that are closely in space but well-separated in time. After the optimization of the number of hits per track by adjusting the module layout, its average time resolution per track (for minimum ionizing particles) can reach to 30 ps at the start of the HL-HLC operation and degrade to 50 ps at the end of that. It will be installed in the very forward region that covers the pseudorapidity of from 2.4 to 4, further augmenting the capabilities of the ITk in that region. With such high precision timing information used to reduce pile-up, the forward objects can be reconstructed with a much higher efficiency, which will certainly improve the physics analyses that are sensitive to the forward region. For example, in the VBF production mode of either single- or double-Higgs boson, one main signature is the outgoing forward jets, whose reconstruction performance can be further translated into sensitivity gains of the physics analysis. In addition, the HGTD is also designed to work as a luminometer. It will provide both online and offline luminosity measurement, aiming to reduce the total uncertainty on the integrated luminosity of HL-LHC compared to Run 2 even with much harsher experimental conditions.

The physics programs of HL-LHC are diverse and vast, see Section 4.1 for more details. Together with other upgraded detectors of the ATLAS, the installation of the HGTD will certainly benefit those physics programs. For example, many SM precision measurements and Higgs properties measurements are already suffering from systematic uncertainties during Run 1 and Run 2 periods, so reduction on systematic uncertainties about the physics objects reconstruction as well as the background modelling will definitely improve the experimental measurements. The HGTD plays an important role in improving the performance of the forward objects reconstruction to the level similar to that in the central region of the ATLAS, which will lead to a reduction on the corresponding systematic uncertainties in the forward region, a new phase space that can not been largely explored during LHC operation. Since the centre-of-mass energy of the HL-LHC collisions is not expected to change much, the focus of many searches for new physics will shift from bump-hunting to looking for discrepancies in the tails of different distributions and other new strategies, such as looking for displaced vertices in long-lived particles searches. Those searches are targeting for new physics just beyond the reach of the collision energy, therefore a more precise understanding of all reconstructed objects in the forward region is required, which can be achieved with the

help of the HGTD. In precision measurements, the precise determination of the luminosity is also crucial. During the first two runs of the LHC, the luminosity uncertainty is already one of the leading uncertainties in the measurements of Higgs couplings. With the precise luminosity information provided by the HGTD, the Higgs properties measurements in the ATLAS will be further improved.

The HGTD will be located in the gap region between the end of the ITK and the end-cap calorimeter, which is currently occupied by the Minimum-Bias Trigger Scintillators. Sealed in hermetic vessel, the detector has the shape of a plate, which will be inserted on both sides at the distance in z of ± 3.5 from the interaction point. The vessel thickness in z is $125\ mm$ (including $50\ mm$ moderator), and it extends from $110\ mm$ to $1000\ mm$ in the radius direction. Unfolded in z direction (see Figure 4.4), it consists of the front cover, two instrumented double-sided layers of disks, inner moderator and back cover, in addition, there are also the outer ring and inner ring for achieving the complete hermeticity. Outside the hermetic vessel, there is also another moderator with the thickness of $50\ mm$ placed behind the HGTD to reduce the back-scattered neutrons created by the end-cap calorimeters and thus protect both the ITk and the HGTD. The instrumented double-sided layers of disk is the main part of the detector and it is consisted of the cooling plate and two working layers that are attached to the plate. The cooling plate is covered with smaller-diameter capillary pipes, aiming to keep the temperature of the vessel at a low level (around $-30^{\circ}C$). The CO_2 is chosen as the coolant due to its high latent heat, low viscosity, capability of mixed states (liquid and vapor) cooling, radiation hardness and low activation. In the working layer, there are the active area extending from $120\ mm$ to $640\ mm$ in the radius direction and the peripheral area extending from $640\ mm$ to $1000\ mm$ in the radius direction, the former is the sensitive part of the detector and directly responds to the particle hits while the latter is the front-end electronics of the detector and handle the large data throughput. From inside to outside, the active area densely covered with $2 \times 4\ cm^2$ modules is divided into three annular regions based on the hit rates, which are inner ring (extending from $120\ mm$ to $230\ mm$), middle ring (extending from $230\ mm$ to $470\ mm$) and outer ring (extending from $470\ mm$ to $640\ mm$). The average number of hits per track for the three annular regions are respectively 2.6, 2.4, and 2.0. The peripheral area is covered with rather large peripheral electronics boards (PEB), which are actually printed circuit board (PCB) mounted with ASICs, connectors and other components.

4.4.2 Hybrid HGTD module

There will be more than eight thousands modules mounted on the active areas of the HGTD. The HGTD module is designed with the structure of sandwich, where two LGAD sensors are located in the middle, and two ALTIROC ASICs are attached to one side while a module FLEX is attached to the other side, see Figure 4.5. With the small size of $2 \times 4\ cm^2$, a

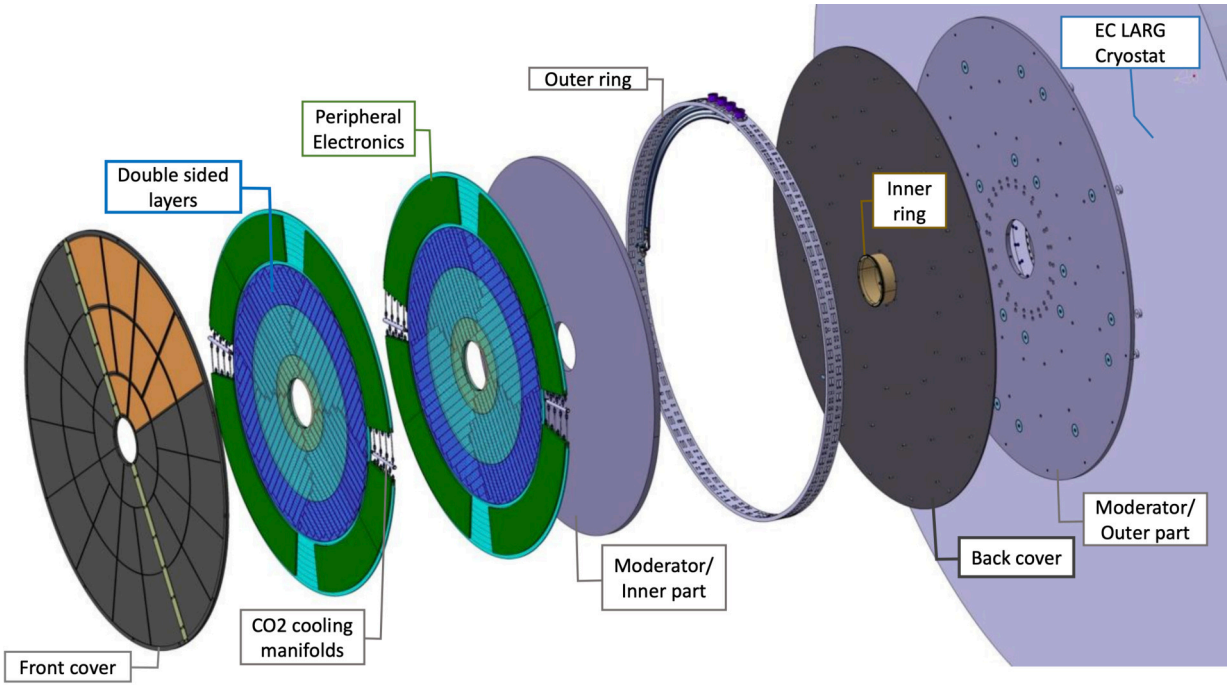


Figure 4.4: A global view of the HGTD, unfolded in z direction [142].

single module has up to 2×225 pixels, and each of those pixels has the size of $1.3 \times 1.3\ mm^2$. When a particle hits the HGTD module, the pad of the LGAD sensor will firstly respond and send out a corresponding analog signal to the ALTIROC ASIC, which will then digitize the analog pulse and process it for further transmission, finally the decoded and serialized data stream gets sent out through the connectors on the module FLEX. There is also being developed a kind of FLEX tail, which is a flexible and flat cable used for attaching to the connectors and transmitting the data out to the PEBs.

4.4.2.1 LGAD sensor

The sensor is based on the novel low gain avalanche detector technology, which was pioneered by the Centro Nacional de Microelectrónica (CNM) Barcelona [143]. Actually this technology is still based on the basic silicon PN junction but with special doping strategy on certain region.

Specifically, the sensor can be viewed as a n-on-p silicon junction with an extra highly-doped p-layer right below the n-p contact region. The doping layer can create a high electric field and trigger the avalanche effect of the charge carriers, thus cause the internal amplification of the current. When the electrons generated by the passing charged track in the depletion layer reach the doping layer (amplification region), new electron-hole pairs will be largely created and thus give rise to a boost of the current. The amplification happening in the doping layer is referred as the gain of the LGAD, which is much larger than that of a standard diode. This large gain is the key factor for LGAD sensor to achieve an excellent time resolution for the minimum ionizing particles. Certainly, the LGAD sensor is sensitive

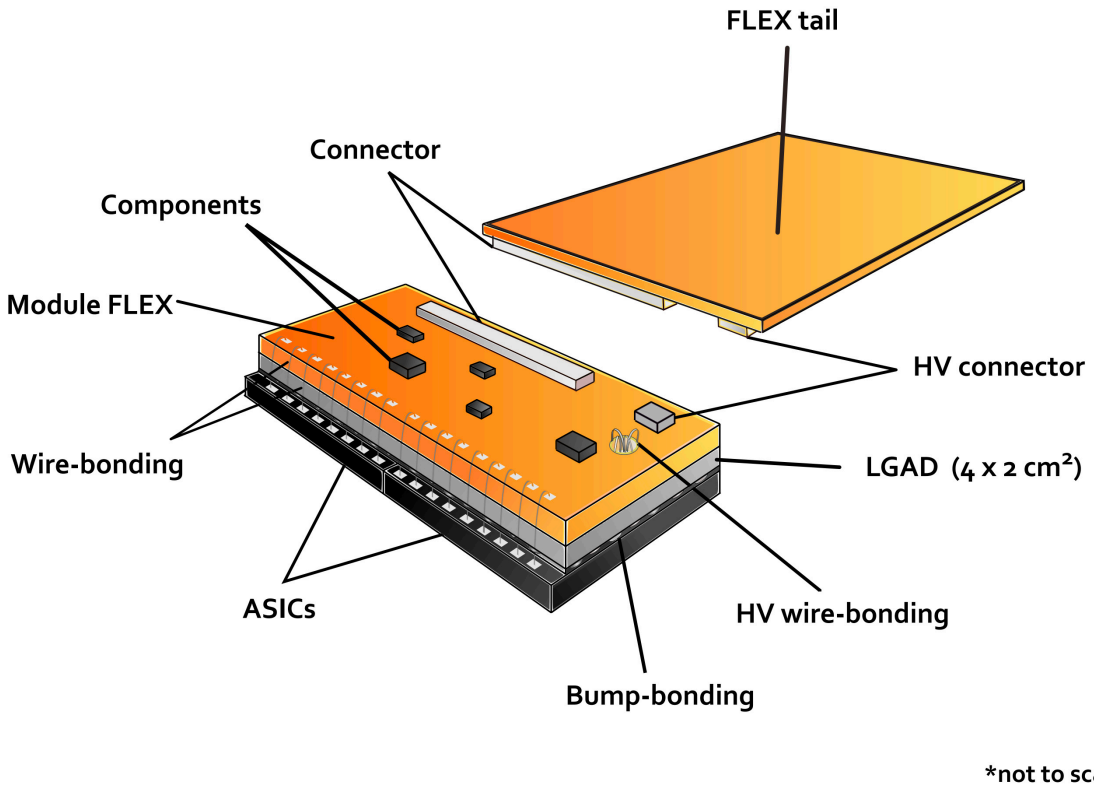


Figure 4.5: View of an HGTD hybrid module equipped with its read-out flex cable tail [142].

to the irradiation and its performance can deteriorate with increasing irradiation dose. For example, a simulation of the signal current in the LGAD sensor shows that [142], before irradiation, the signal can last for about $1\ ns$ with the rise time of about $500\ ps$ and the collected charge can reach up to $12fC$, while after the irradiation of the $1\ MeV$ neutron equivalent flux of $2.5 \times 10^{15}\ N_{eq}cm^{-2}$, the both the rise time and the signal duration become shorter so that the collected charge decreases to just $4fC$. In order to mitigate the adverse effect of the irradiation like the decrease of the charge yield, the bias voltage of the LGAD will be increased accordingly with the gradually accumulated irradiation dose, and it can be set up to $750V$ at the end of the lifetime, in addition, the detector will be operated at low temperature (around $-30^{\circ}C$).

In order to make the performance as high as possible, the LGAD has been extensively designed and tested to determine some key parameters. By tuning the doping levels of the gain layer, the inter-pad gaps, the pad size and the active thickness, many different kinds of LGADs have been produced by different companies from across the world, and the characterization of those LGADs are being intensively performed after the exposure to different irradiation sources (protons, neutrons and X-rays) to the expected maximum radiation levels. For the determination of the active thickness of the LGAD, on one hand, the thinner it is, the smaller the electronics jitter will be, since the jitter is in inverse proportion to the rising rate of the output voltage (dV/dt), which is also inversely proportional to the thickness of the sensor, on the other hand, the jitter has a positive linear correlation with the

detector pad capacitance, which means the thicker it is, the smaller the electronics jitter will be. Therefore the final selection of the active thickness is a compromise of the two effects. The active thickness of $50\ \mu m$ gives a rather optimal performance, so it has been chosen to be as the baseline active thickness while the total thickness is $250\ \mu m$. For the determination of the pad size, it is also a compromise of smaller pads and larger pads. Smaller pads give smaller detector pad capacitance, which lead to low electronics jitter, and also smaller pads can give a lower occupancy of the detector. If adopting larger pads, there will be a better geometric coverage with large fill factors and less power consumption required for the ASIC. Finally, a rather optimal pad size of $1.3 \times 1.3\ mm^2$ has been chosen as the baseline.

The time resolution is the key indicator that evaluates the performance of the HGTD LGAD sensor. There are three main effects that determine the time resolution, and they are respectively the time walk effect, the jitter from electronics noise and the Landau fluctuations effect. Firstly, for the time walk effect, it describes the fact that the determination on the arrival time of several signal pulses with different amplitudes by threshold setting may be different even through all those signal pulses arrive at the same time. The time resolution contributed from the time walk effect can be written as

$$\sigma_{TimeWalk} = [V_{th}/\frac{S}{t_{rise}}]_{RMS}, \tag{4.1}$$

where the V_{th} is the threshold voltage set for the discriminator of the ALTIROC ASIC, S is the signal that is proportional to the gain, and t_{rise} is the rise time. As can be seen, a smaller rise time can mitigate the time effect. To achieve a smaller rise time, the active thickness of the sensor should be thin. However, this can not totally solve the adverse impact from the time walk effect, usually the time walk can be corrected by using some time reconstruction algorithms like constant-fraction discrimination (CFD) and time-over-threshold (TOT) corrections. For the electronics noise jitter, its contribution to the time resolution can be written as

$$\sigma_{Jitter} = \frac{N}{(dV/dt)} \approx \frac{t_{rise}}{(S/N)}, \tag{4.2}$$

where N is the noise, dV/dt is the voltage rising rate at the output of amplifier. As can be seen, a higher signal-to-noise ratio (SNR) can reduce the time resolution contribution from electronics jitter, which requires a larger gain of the detector. Clearly both the time walk and the electronics noise jitter are caused by the imperfection of the front-end electronics, so the selection of the readout electronics is extremely important for achieving a good time resolution. For the Landau fluctuations effect, it describes the fact that the responses to different particles that hit the sensor at the same time varies because of the non-uniform charge deposition when the particles pass through the depletion layer of the LGAD. Therefore the time-of-arrival (TOA) that will be immediately determined by the ALTIROC ASIC will also

fluctuate. To mitigate this effect, the thickness of the sensor should be made thinner and a threshold value of the ALTIROC discriminator should be carefully selected.

4.4.2.2 ALTIROC ASICs

The ALTIROC ASIC is the dedicated chip that is being developed for the front-end electronics of the HGTD. It will be bump-bonded to the LGAD sensor for processing the analog pulses created by the particles hits and sending them out to the peripheral electronics of the HGTD, described in Section 4.4.3. In order to better handle the total ionizing irradiation and satisfy the quite lower jitter requirement, the complementary metal-oxide-semiconductor (CMOS) $130\ nm$ technology from Taiwan semiconductor manufacturing company (TSMC) has been chosen to produce this ASIC. Concerning the geometrical parameters of the ALTIROC, it will be designed with a total area of $19.9\ mm \times 21.7\ mm$, where there will be a 15×15 pixel matrix area with the size of $19.5\ mm \times 19.5\ mm$ and an additional area for holding the peripheral logics of the chip. In general, the ALTIROC is structured in this way, the on-pixel logics are repeated by 225 times on the pixel matrix area, and on the peripheral area, there are the end-of-column (EOC) logics dealing with each column of 15 pixels, the control unit configuring the working parameters of the chip, the analog part for monitoring temperature and leakage current and providing clean clocks with low jitter. the trigger data processing unit (TDPU) handling the timing data from the buffers of the pixels, and the luminosity processing unit (LPU) handling the luminosity data also from the corresponding luminosity logics of the pixels.

In terms of the on-pixel logic, it starts from the analog electronics that have direct ohmic contact with the LGAD pixel. More specifically, the analog signal from the LGAD sensor firstly goes to the preamplifier of the ALTIROC pixel and then gets amplified. The bandwidth of the preamplifier is designed to about $1\ GHz$ after considering the fact that the LGAD pulse duration is approximately $1\ ns$. Then the amplified signal will be fed into the discriminator, which is used to shape the analog pulse to square pulse in a certain way that can be set by adjusting parameters of the discriminator. The leading edge of the square pulse is referred as the TOA of the signal while the duration of the square pulse is referred as the TOT of the signal. Those are exactly the timing information that we are targeting at. For the measurement of the TOA, the time interval between the leading edge of the square pulse and the next LHC clock edge will be recorded, with this time interval, the TOA is actually known to us by just going back by this certain amount of time from the LHC clock edge. The time interval will be measured with the first time-to-digital converter (TDC) by using the leading edge of the square pulse as the start signal and the the next LHC clock edge as the stop signal. The TDC adopts the Vernier delay line structure that is accurate to a quantization step of $20\ ps$. This kind of structure can significantly reduce the power consumption since it doesn't need to be in the full operation mode when no hits happen.

In addition, due to the fact that all the hits of the collisions are centered on the bunch crossing with the time dispersion of about 300 ps, the measurement window for TOA is set to be 2.5 ns centered on the bunch crossing. For the measurement of the TOT, similarly the leading edge and the falling edge of the discriminator output square pulse are fed into the second TDC respectively as the start and stop signals. Compared with the first one, the second TDC has a rather coarse quantization step (40 ps) because the resolution on the TOT measurement can be a bit larger than that on the TOA measurement. The purpose of measuring TOT is to have an estimation of the signal amplitude, which can be used offline to correct the TOA affected by the time walk effect. Finally, both the TOA and TOT information will flow into the digital frontend of the pixel and be temporarily stored in the buffer until the reception of the L0/L1 trigger.

In terms of the logics on the peripheral area, it has to be able to handle the data flow processing, configuration and monitoring between the pixel matrix area and its peripheral area and also between the chip itself and the electronics outside the module. The EOC logic will firstly handle the timing information of the triggered hits from all the 15 pixels of this specific column, at the meantime, it will also do the summation of the luminosity hits that happened in this column. Actually, the luminosity hits are summed in two different time windows that are all centered on the bunch crossing, one has a fixed width of 3.125 ns, the other one has a larger width that can be configured through the control unit in the steps of 3.125 ns. After the EOC logic, the timing data will go to the TDPU for data formatting and then get serialized by its serializer, which can be configured to work at the speed of 320 $Mbit/s$, 640 $Mbit/s$, 1.28 $Gbit/s$ through the control unit to maximize the use of the bandwidth. The serialized data will be further transmitted to e-link of the lpGBT through the FLEX cable that connects the module and the PEBs. In a similar way, the luminosity data after summation will go to the LPU for data formatting and then get serialized by its serializer, which in this case can only be working at the speed of 640 $Mbit/s$. For the control logic, it is consisted of three parts, the register array for storing the configuration values for the chip and also key parameters reflecting the working status of the chip, the slow control logic for receiving slow control parameters from other chips through the I^2C bus, and the fast command unit for receiving the fast commands (such as clock, L1 trigger, BCID, etc) from the lpGBT. In addition, there are also some analog logics

4.4.2.3 Module assembly and loading

As shown in Figure 4.5, the hybrid HGTD module is consisted of three parts, the module FLEX, the two ALTIROCs and the two LGAD sensors, which have to be put together to be able to get the final functioning module. The idea is to firstly bond the ALTIROCs and LGAD sensors together to make the bare module, then connect the bare module and the module FLEX together. The process of making the bare module is called flip-chip bump

bonding, which generally contains two steps. The first step is to make sure that the under bump metallization (UBM) is properly deposited on the sensor wafer before dicing, and also both the UBM and the solder bumps are deposited on the ALTIROC wafer. The second step is to do the flip-chipping, which aims to create the robust electrical contact between the LGAD pads and the ALTIROC pixels after being aligned, heated and compressed. A crucial point that has to be mentioned in flip-chipping process is the alignment of the sensors and ASICs. Only with the well-aligned sensors and ASICs, can a reliable and operative module be guaranteed. With the bare module well produced, it's time to make the full HGTD module by gluing the bare module and the module FLEX together. This gluing process is certainly not as easy as how we glue things in daily life, and it requires investigations and tests, such as the selection of the glue types, the amount of glues used on the contact surface, the gluing patterns, the procedures of the gluing process, etc. In addition to the gluing that attaches the bare module to the module FLEX, there are also wire bonding connectivities between them, which are actually the wire bondings for ASIC signals and the high voltage.

There are many passive components mounted on the module FLEX, among which two mini-connectors are very important. One connector is used for the signal transmission between the module and the PEBs while the other one is used for the high voltage power supply. Dedicated FLEX cables are being designed for the two mini-connectors, and they will be attached to the connectors on the module FLEX and the same connectors used on the PEBs. For the design of the FLEX cables, some key points have to be taken into account. Due to the very limited space of the HGTD vessel in the z direction and also the fact that there will up to 19 modules in one readout row, the thickness of the FLEX cable must be less than $220\ \mu m$. Since distances from the module to the PEBs differs, the FLEX cables have to be designed with various length, which certainly brings more complexities to the design. Either the serialized data stream with the high speed of up to $1.28\ Gbit/s$, the fast commands from lpGBTs with the speed of $320\ Mbit/s$, or the clocks distributed to the ALTIROCs all require efficient and robust transmission, so the FLEX cables have to be designed with high-performance transmission line of properly-adjusted impedance. There should be high voltage delivered to the frontend as the bias voltage of the sensor (up to $750\ V$), therefore the high voltage FLEX cables require excellent insulation. Another requirement is about the resistance of the transmission line of the FLEX cables, to make the voltage drop of the ALTIROC power supplies and the power dissipation as low as possible, the resistance is required to be less than $300\ m\Omega$ for the longest power supply line of the FLEX cable.

At the end, the hybrid HGTD modules have to be loaded on the support structure to make the detector instrumented. The dedicated support units are designed for holding the HGTD modules and ensuring the exact position of each module. The support units will be screw on the cooling plate, and the modules are then inserted in those slots of the support units, to make them more firmly attached, the modules will be further glued on each side

of the slots. The baseline material of the support units is chosen to be carbon fiber. The module layouts in the inner ring, middle ring and outer ring are different due to the hit rates in different rings, the more inner the module is located, the denser the layout will be. In the inner ring, the modules are aligned in a step of 25.5 *mm* in a given row, and the corresponding overlap between top side and bottom side modules is up to 70%. The step is 28.4 *mm* and the overlap is 54% in the middle ring, while the step is 34.5 *mm* and the overlap is 20% in the outer ring. To guarantee good thermal contact between the modules and the cooling plate, a conductive grease will be used to efficiently conduct the heat from the modules to the cooling plate. A rather comprehensive thermal simulation of the system has been performed by taking into account many key factors, such as the best understanding of the thermal contact of all the materials, the expected power dissipation of the sensors, the working temperature and the accumulated irradiation dose received by the system. The simulation results show that, even with the highest power dissipation and large temperature variation of up to $25^{\circ}C$, there is no thermal runaway observed, which reassures us that the system can have safe operation under all challenging thermal conditions.

4.4.3 Peripheral electronics of the HGTD

4.4.3.1 Overview

The peripheral electronics of the HGTD serves as a bridge that connects the HGTD modules and the off-detector systems like the data acquisition (DAQ) system, the luminosity system and the detector control system (DCS). Figure 4.6 shows how those components of the peripheral electronics are connected with each other and how the signals flow upwards and downwards. As can be seen from the diagram, the peripheral electronics can handle the diverse connectivities with huge complexities, including the timing data stream with the speed of 320 – 1280 *Mbit/s*, the luminosity data stream with the fixed speed of 640 *Mbit/s*, the fast commands and trigger stream in with speed of 320 *Mbit/s*, the clocks distributed to the front-end electronics, the monitoring of the temperature and current in the frontend, the slow control, the low voltage and high voltage distributions. There are four main ASICs that will be used for peripheral electronics, which are the lpGBT for the data processing, the bPOL12V for the low voltage regulation, the MUX64 for picking up different monitoring signals and the VTRx+ for the electro-optical conversions. Those ASICs have to be put together on the PEB in a specific and optimal way such that the peripheral electronics can function correctly and reliably in the harsh environment during its lifetime. In addition to those ASICs, of course, there will be also some passive components mounted on the PEB, among which the connectors take rather large space on the the surface of the board.

There are two main data streams that the peripheral electronics should handle, one is the uplink timing data and the other one is the uplink luminosity data. For the timing data,

starting from the frontend, each serialized timing data from each ALTIROC of the module flow through the differential transmission lines of the FLEX cable, and gathers at the lpGBT electrical ports that have the bandwidth of up to $1.28\ Gbit/s$. Depending on the data rate, one single electrical port of the lpGBT can handle from 1 to 4 serialized timing data links. There are 7 such electrical ports on one lpGBT chip, which means many modules can be attached to one lpGBT. After collecting the serialized timing data at the electrical ports, the lpGBT working in the transceiver mode (DAQ lpGBT) will further pack the incoming data streams into an even larger dataframe, which will be successively scrambled, encoded, interleaved and finally serialized to a high-speed link with the data rate of $10.24\ Gbit/s$. Then the serialized data will be converted to the optical signal by the VTRx+, which will be transmitted out to the DAQ system through optical fiber. For the luminosity data, it goes through a similar data path to the timing data except that a dedicated lpGBT working in the simplex transmitter mode (lumi lpGBT) will be used.

In terms of the power supplies, there are basically two kinds of powers delivered to the peripheral electronics of the HGTD. One is the high voltage power supply, which will be further delivered to the modules as the bias voltage of the LGAD sensors. The high voltage can be increased up to $750V$ at the end of the HL-LHC operation. Actually, the high voltage just passes through the peripheral electronics, more specifically, there will be some dedicated wires on the PEBs where high voltage power can flow and it doesn't get changed. The other one is the low voltage power supply, which is about $10\ V$ when it arrives at the peripheral electronics. Then it will get regulated by the DC-DC converter called bPOL12V to lower voltages, $1.2\ V$ for the power supplies of the lpGBTs, VTRx+ and the front-end ASICs, $2.5\ V$ for the power supplies of the lpGBTs, VTRx+ and MUX64. Since those DC-DC converters will be designed to be enabled by the lpGBT, a converter should avoid providing power supplies to the lpGBT that enables this specific converter. In terms of the monitoring, it aims to monitor five voltages of each module, including the two voltages from the temperature sensors of two ALTIROCs, the analog and digital power supply voltages received at the ALTIROCs, and the voltage related to the current return of the module. The number of modules that one PEB can handle is up to 55, it is not reasonable to route all the voltages out through the PEB to the DAQ system. Therefore, the 64-to-1 multiplexer is used on the PEB to circularly read out each of them by using strobe signal coming from the lpGBT.

In terms of the control of the front-end modules, there are also two types. One is the slow control performed by the I^2C bus from the lpGBT to the ALTIROC, and it is used to talk with the configuration register array of the ALTIROC and thus change or acquire the status of the ASIC. The original configuration information comes from the DCS side, and it gets embedded in the high-speed downlink data stream, which can be transmitted through the optical fiber from the FELIX card to the VTRx+ mounted on the PEB and then gets

processed by the lpGBT, finally the control information gets extracted out and distributed to the ALTIROCs. The other one is the fast command performed by the lpGBT electrical downlinks at the speed of 320 $Mbit/s$, and it aims to deliver those fast signals like clock, L0/L1 trigger, bunch crossing identification (BCID) and other configuration parameters that require rapid adjustment. Similar to the slow control, the original fast control information also comes from the DCS system, and finally delivered to the ALTIROCs in the same way.

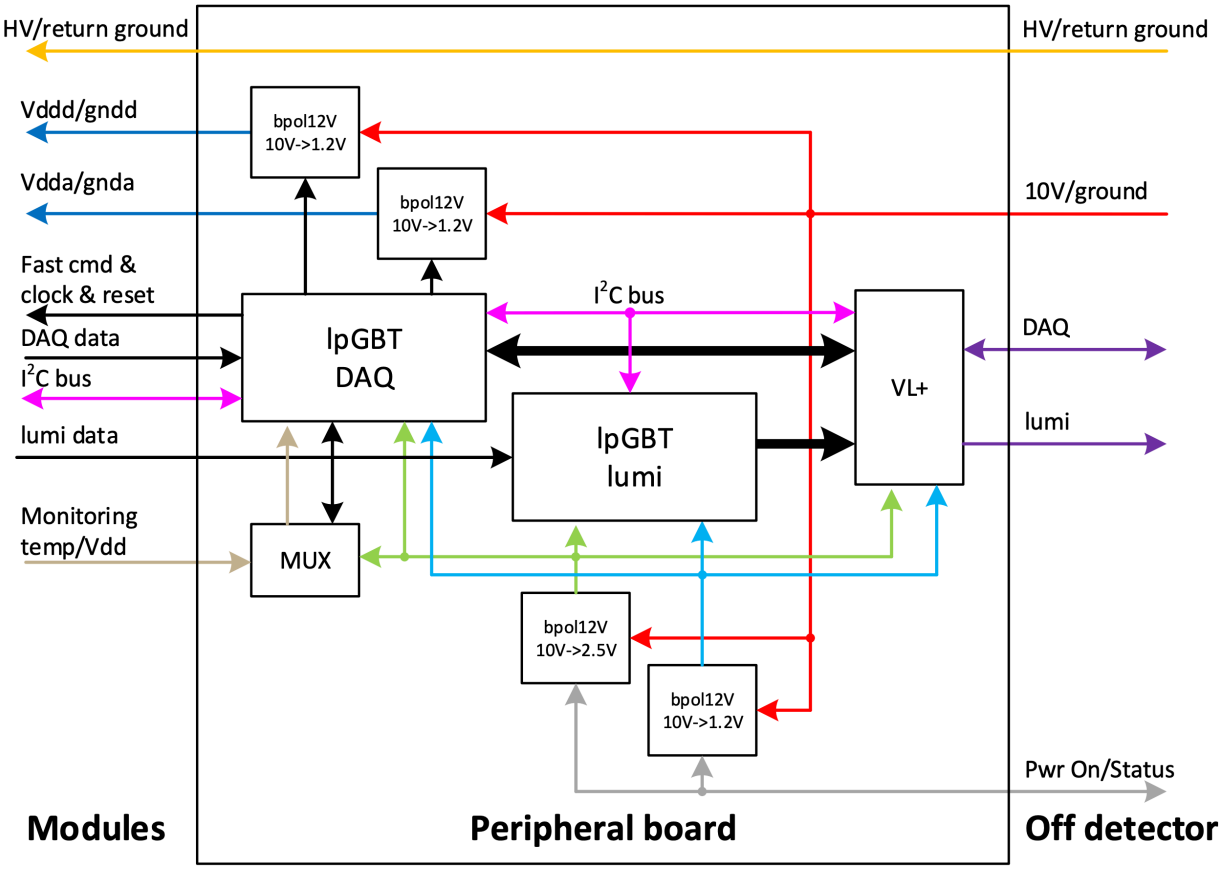


Figure 4.6: Block diagram of the peripheral electronics of the HGTD [142].

4.4.3.2 lpGBT

The lpGBT ASIC (Low Power GigaBit Transceiver) [144] is a 65nm CMOS radiation-tolerant serializer/deserializer device developed by CERN for the HL-LHC detectors. It is a very powerful ASIC with diverse functionalities, such as data aggregating and serialization, clock distribution, configuration of other devices via many ways (I^2C, 320 $Mbit/s$ electrical downlink, and even GPIO). In practice, the lpGBT is used together with the electro-optical converter VTRx+. Figure 4.7 shows the schematic of the lpGBT architecture.

Two important parts of lpGBT would be the uplink logic and downlink logic [145]. For the uplink logic, starting from the seven electrical ports with the bandwidth of up to 1.28 $Gbit/s$, one single port can receive 32 bits from one 1.28 $Gbit/s$ serialized link, or two 640 $Mbit/s$ serialized links, or four 320 $Mbit/s$ serialized links at the rate of 40 MHz, therefore

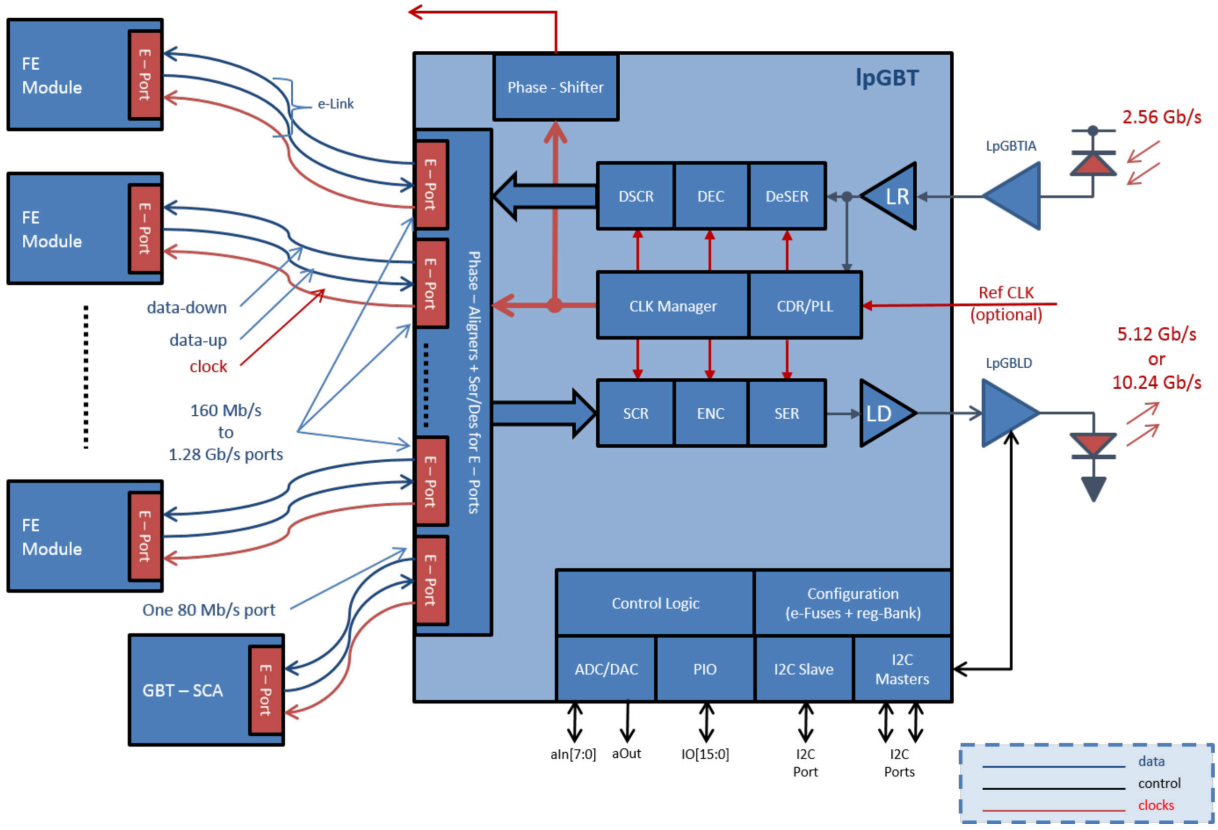


Figure 4.7: Schematic of the lpGBT architecture [145].

seven ports together can receive 224 bits at the rate of 40 MHz. Then those 224 data bits together with header bits, forward error correction (FEC) bits and other bits, will be packed into a 256-bit dataframe, which will be successively scrambled, encoded, interleaved and finally serialized to a high-speed data stream with the rate of 10.24 $Gbit/s$. Another thing that needs to mention is that, actually before the electrical port receive the 32-bit data, there is a process of deserialization, which just simply samples the serialized data and fill it into the 32-bit frame. The phase aligner circuit can make sure that the serialized data can always be sampled at the optimal phase, which allows the different data streams from FLEX cables to be connected to the same lpGBT even the length of the FLEX cables are different. For the downlink logic, starting from the 2.56 $Gbit/s$ high-speed data stream, it will be firstly deserialized to a 64-bit dataframe (also including data bits, header bits, FEC bits and other bits) at the rate of 40 MHz, which will be successively deinterleaved, decoded, and finally descrambled. The 32 data bits from the descrambled dataframe will be further split into four groups of 8 bits, which are then pushed into four electrical ports where each 8-bit frame is serialized into 320 $Mbit/s$ data stream. Those 320 $Mbit/s$ data streams are sent to the front-end electronics for delivering fast commands. On each electrical port, the 320 $Mbit/s$ data stream can be mirrored to 3 another exactly same ones, which sums up to 16 electrical links for fast commands. Therefore, it gives the limitation of not more than 16 ALTIROCs connected to the same lpGBT, which defines the minimum number of lpGBTs

required by the peripheral electronics.

In addition to uplink and downlink logics, there are also many other important logics that extend the functionality of the lpGBT to a even higher level. Those logics are briefly described as follows, including the phase shifter for providing four programmable and high-resolution differential clocks, the clock generator for generating up to 29 differential clocks with configurable frequencies, the general purpose I/O (GPIO) block for providing 16 highly-configurable GPIO pins, the I^2C master block of 3 I^2C buses for controlling other devices as I^2C slaves, the I^2C slave block with just one I^2C bus for receiving configuration from other I^2C master devices, the high-speed line driver for further driving the uplink high-speed serialized data stream that is about to leave the lpGBT, the high-speed equalizer for equalizing the incoming high-speed $2.56\ Gbit/s$ downlink data stream before it's further processed by the lpGBT, the power-up state machine (PUSM) for handling the operation of the lpGBT in a general, the configuration block of e-fuses and registers array, and the analog peripherals of 8 analog-to-digital converter (ADC) pins and 1 digital-to-analog converter (DAC).

It's very important to configure such a complex chip in a robust and smooth way. Actually, there are several ways to configure the lpGBT partially or completely. Firstly, through the I^2C slave port, the lpGBT can be configured by other I^2C master ASICs, such as another lpGBT with one of its three I^2C master ports as the master. Another way is using the e-fuses array of the lpGBT, which should be programmed in advance and can be loaded automatically after powering up. However, due to very limited data capacity of the e-fuses array, only the information of some key parameters can be burned over there, also this kind of programming is unerasable and can be done just once. Last but not least, the very commonly-used lpGBT configuration strategy is the cascade configuration, where the master lpGBT firstly gets configured by the internal bits embedded in its downlink stream and then broadcast the external bits also embedded in the same downlink stream to the slave lpGBT through the dedicated external pins, finally the slave lpGBT will be also properly configured. Actually, from the point view of the lpGBT, all the three ways of configuration are essentially the same, since at the end the procedures will all boil down to the update of the register array, from where the lpGBT adjusts its parameters and thus the changes of working status.

4.4.3.3 DC-DC converter bPOL12V

The bPOL12V is one of the ASICs developed for the CERN DCDC converter project [146], aiming to efficiently regulate the voltage of about $10\ V$ to lower voltages under the irradiation environment. It will supply the $1.2\ V$ voltage required by the power supplies of the ALTIROCs and lpGBTs, and $2.5\ V$ voltage required by the laser driver of the VTRx+, the e-fusing of the lpGBT and the MUX64 power supply. On the peripheral electronics, the low voltage power supply will be performed by this ASIC with a single stage conversion (i.e.

from the $10\ V$ input voltage directly to all the target voltages) instead of the dual stage conversion that potentially has higher efficiency. The motivation of adopting the single stage conversion is entirely due to the extremely-limited surface on the PEBs.

The input voltage of the bPOL12V can varies from $5.5\ V$ to $12\ V$, and its output current can reach up to $4\ A$. Its output voltage can be also precisely set by adjusting the values of the related resistors. It's found that, in order to reduce the possible switching transients, the input voltage should not be over $10\ V$ when working near the maximum current. Also some measurements show that, under the temperature of $-30\ ^{\circ}C$ and the output current of $3\ A$, the highest efficiency of up to 72% can be achieved. The requirement on its irradiation tolerance is being able to withstand up to $2\ MGy$ TID dose and $2.5 \times 10^{15}\ n_{eq}cm^{-2}$ neutron equivalent dose. There is already a PCB layout based on the ASIC by the bPOL12V design group, but it has a quite large profile of not only big footprint but also big height due to the usage of the large size inductor. Therefore, the large PCB layout can not fulfill the requirement of the PEB where we there is very limited space available, especially along the z direction. A customized PCB layout is being developed to minimize the profile of the device, where several types of inductors with smaller size and also shielding cages are being tested.

The power-on sequence is briefly presented as follows. The bPOL12Vs will not only supply the ASICs on the PEB itself but also the front-end ASICs. At the very beginning of powering up, all the bPOL12Vs are firstly turned on by an external $1-V$ enable signal from the DCS system, then all the other kinds of ASICs on the PEB are turned on. The status of the bPOL12Vs will be fed back to the DCS system via external electric cables and the GPIO of the DAQ lpGBTs that they don't supply. This can help us to recognize where the possible failures come from, the lpGBTs or the bPOL12Vs themselves. Once the lpGBTs and VTRX+ are powered on, those bPOL12Vs supplying the ALTIROCs will be activated by the enable signals coming from the GPIO pins of the lpGBTs, and also the status of those bPOL12Vs will be sent to the lpGBTs. Those enable signals from the lpGBTs are interpreted from the DCS bits embedded in the lpGBT downlink high-speed data stream. For the power supplies of the ALTIROCs, they are divided into two groups, the analog power supply conducted by some bPOL12Vs, and the digital power supply conducted by different bPOL12Vs. The analog power supply of up to 6 ALTIROCs (that is, 3 modules) will be handled by one bPOL12V, while digital power supply is handled in the exact same way. For the power consumption of the ALTIROCs, each ALTIROC can at most consume $1.2\ W$, including $0.5\ W$ for analog part and $0.7\ W$ for digital part. In practice, the power consumption will really depend on the location of the ALTIROC that represents the average number of hits, the more inner it is, the larger power it will consume. In addition, it's also estimated that the power consumption of the DAQ lpGBT will not exceed $0.55\ W$ while it's $0.45\ W$ for lumi lpGBT due to its downlink path being closed.

4.4.3.4 VTRx+, MUX64 and connectors

The VTRx+ is the output of the versatile-link-plus project [147] targeting at the Phase-2 upgrade of the HL-LHC experiments. It is designed to be a bi-directional electro-optical converter with good irradiation-hard performance, and it's expected to withstand up to $1\ MGy$ TID dose and $1 \times 10^{15}\ n_{eq}cm^{-2}$ neutron equivalent dose, which are larger than the doses received at its installation location (beyond $85\ cm$ in the radius direction). Its uplink is able to handle the data transmission between the on-detector electronics and the off-detector electronics at the rate of up to $10.24\ Gbit/s$ while its downlink can deal with the configuration data stream from the DCS system to the on-detector electronics at the rate of $2.56\ Gbit/s$. For each VTRx+, There are four channels for the uplink transmission and just one channel for receiving downlink DCS data stream. The asymmetrical design of the uplink and downlink data rate is based on the fact that the bandwidth required by the DCS system is always way lower than the bandwidth of the upstream data from the frontend across all the HL-LHC detectors. Each DAQ lpGBT needs to take one uplink channel and one downlink channel of the VTRx+ while each lumi lpGBT just needs to take one uplink channel. The VTRx+ is also designed to be low power consumption, low material contribution, small profile and easy deployment. The VTRx+ is packaged as a module with a footprint of $20\ mm \times 10\ mm$, which is pluggable through electrical connectors. In addition, a optical fiber pigtail extends out of module and terminated with 12-core MT type optical connector.

The MUX64 is a fully-customized 64-to-1 multiplexer ASIC designed with CMOS $130\ nm$ technology, and it has been produced in December 2019 and characterized since then by a variety of ways like neutron beam test, x-ray irradiation test and high-low temperature cycling test. So far, the results from those tests are very promising. The MUX64 will be used on the PEB to handle the signals to be monitored from the front-end modules. There are five signals to be monitored for each module, which are the operating temperatures of the two LGAD, the leakage current of the whole module, the received analog and digital voltages received by the module. A monitoring structure based on the MUX64, the GPIOs and ADC of the DAQ lpGBTs has been proposed. With this structure, one lpGBT is able to monitor up to 7 modules, which means 7 groups of 12 modules are respectively connected to 7 MUX64 on the PEB through FLEX cables. The 7 outputs of the 7 MUX64 are attached to 7 ADC input pins of the lpGBT and processed by the corresponding ADC. The 6 channel selection pins of each MUX64 share the 6 GPIO pins of the lpGBT, which are controlled by the DCS downlink bits and collectively select a certain channel every time for all the 7 MUX64. In practice, how the modules are grouped can be adjusted according to the actual situation during the design of the PEB. The ADC of the lpGBT is of great importance in the entire monitoring chain. It is a differential 10-bit ADC with a large input dynamic range of from -VREF to VREF, where VREF is the reference voltage for the ADC circuit and can be provided either internally or externally. In addition, the ADC has a gain circuit

implemented and a gain value (2x, 8x, 16x and 32x) can be configured to amplify smaller input voltages.

Different kinds of connectors are used on the PEBs to interface with both on-detector and off-detector electronics for diverse connectivities. One of the main considerations on the connectors selection is the size of them because of the strict space constraint on the PEBs. Therefore, both the height and footprint of the candidate connectors should be taken into account. To handle the FLEX cables of a readout row from the modules to PEB, two types of high-density mini-connectors has been chosen as the baseline connectors, one is a 71-pin connector with the footprint of $22.8\ mm.5\ mm$ and the height $1.0\ mm$ for the signals transmission and low voltage power supply while the other one is a 13-pin connector with the footprint of $5.4\ mm.5\ mm$ and the height $1.0\ mm$ for the high voltage power supply. They also have good voltage insulation between two adjacent connector pins, which allows the high voltage connector to hold large bias voltages for the LGAD. The high voltages delivered to the HGTD vessel pass through the high voltage feed-through connectors on the out ring, and a commercial 56-pin connector has been selected as the baseline. Similarly, there will need the feed-through connectors for the $10\ V$ low voltage power supply, and its number of pins is way smaller than that of the high voltage feed-through connector since the DCDC converters on PEB can share the low voltage power supply while each HGTD module has to be supplied with individual high voltage. In addition, there are also feed-through connectors for the optical links, to match with the 12-core MT connector of the VTRx+ pigtail, the same type of connector will be used. The pigtails can be extended with optical patch cables to be able to reach the feed-through connectors on the out ring.

4.4.3.5 Peripheral electronics boards

The peripheral electronics is implemented at the peripheral area, which will be covered with peripheral electronics boards (PEB) mounted with a large amount of components. The physical space that can be used by the peripheral electronics is extremely limited. The total thickness of the HGTD vessel is only allowed to be $75\ mm$, so the space available for the four layers of PEBs along the z direction is just $9\ mm$ with a margin of $1\ mm$. In the radial direction, it is constrained within the range of $660\ mm$ to $920\ mm$. Within such narrow and restricted space, all those ASICs, connectors and mechanical supports have to be properly arranged to construct a operative and reliable peripheral electronics system, which poses great challenges to the design of the PEBs.

The peripheral electronics is split up into five types of PEBs per quadrant, and a similar number of modules is connected to each PEB. In principle, due to the location difference between the modules on the front and on the back, the five PEBs on the front is also different from the five PEBs on the back. After some optimization of the module location, the total number of the PEB types for both the front and back is reduced to be 6, meanwhile the

number of the FLEX cable length is also significantly decreased. To make the lpGBTs connected with similar amount of modules, the 21 readout rows have to be combined in this specific way, 1 – 3, 4 – 7, 8 – 14, 15 – 18, and 19 – 21, see Figure 4.8. This kind of combination can lead to the reduction in the number of ASICs used on the PEBs and thus give more room to the mini-connectors that are as many as the modules. After the determination of the connection strategy between modules and PEBs, many key points are thus clear, such as the exact number of modules connected to each type of PEBs and rather precise numbers of the key ASICs (DAQ lpGBT, lumi lpGBT, VTRx+, DCDC converters and MUX64) used on each type of PEBs, which helps the designer to have a better understanding of the PEB and think some key points (like testing some parameter settings and validating technologies that will be used in the production stage) even ahead of the PEB design. Taking the most complex PEB1 (PEB 1F) as an example, there will be up to 55 modules attached to it, which means 110 mini-connectors are required. Therefore, the distance between two adjacent connectors should be as small as possible to make more room to other components. This distance is being investigated through the modular PEB and the prototype. The crowded surface of the PEB1 will also make the wire-routing become very challenging.

In the actual PEB design, there are several considerations that need to be taken into account. To reduce the consequences of possible failure components on the PEBs, the exact electrical connectivities between modules and the components have to be carefully designed in such a way that the number of modules affected is as small as possible. To achieve that, firstly, the modules that share the same DAQ lpGBT should be spread as far as possible, and it should also be the same lpGBT that distribute their clocks and fast commands, control their DCDC converters, deliver their DCS configuration and handle their monitorings. For the same reason, those modules sharing the same DAQ lpGBT should also share lumi lpGBT, and the lumi lpGBT should be as a cascading lpGBT controlled by that DAQ lpGBT. Another point is that, the DCDC converters and lpGBTs are required to have a proper radial layout so that their power dissipation can preheat the CO_2 cooling plate in a more efficient way.

4.4.4 DAQ, DCS and luminosity

The HGTD DAQ system will be embedded into the the ATLAS DAQ common readout. It can be divided in two parts, the on-detector electronics located in the experimental hall, and the off-detector part located in the USA15 counting room. The interface between the on-detector and off-detector electronics is implemented via the versatile link (based on the lpGBT and VTRx+), where the luminosity data stream and the main data stream (including the timing data and DCS data) are passed through. For the off-detector electronics, it can be taken as a system based on the FELIX board [148], which receives timing and luminosity data from the on-detector electronics and sends them to the data handler through the multi

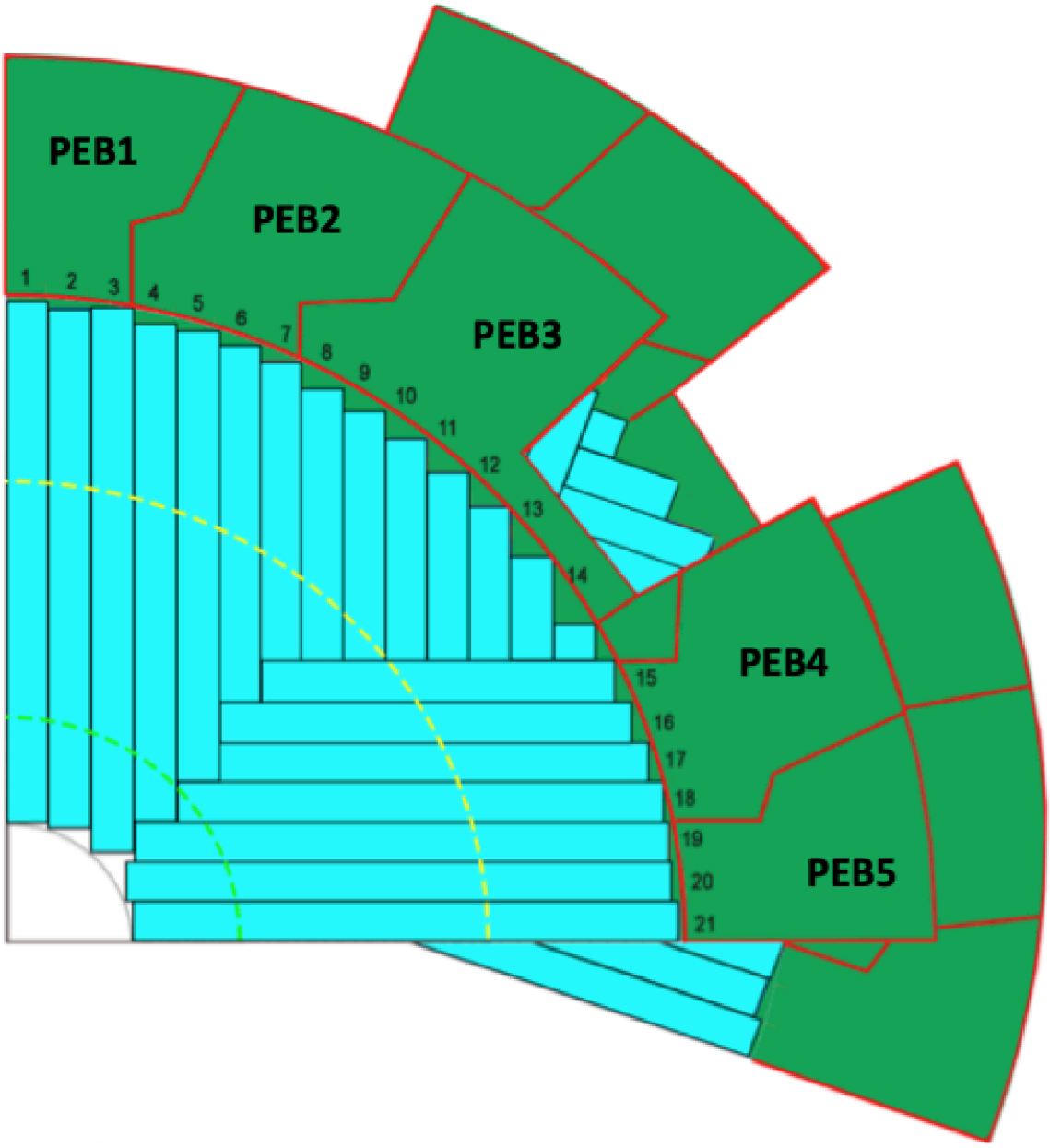


Figure 4.8: One quadrant of the HGTD front and back disk is shown [142].

gigabit network. In addition to that, the FELIX also interacts with the timing, trigger and control (TTC) system through the local trigger interface (LTI), and with the DCS system for control and monitoring. After receiving the event data, the data handler will decode them for further event building within the ATLAS TDAQ infrastructure. Then the decoded data will be sent to the data flow for event filter. The average event size is expected to be $250\ kB$ with a range of $150 - 350\ kB$. An application HGTD controller is developed to handle the configuration of the modules and lpGBTs, and the HGTD calibration. By scanning and fine-tuning different parameters, the calibration can push the detector to work in a optimal condition where the measurements are experimentally correct. During the calibration procedure, up to 3.4 millions of ALTIROC channels have to be read, which gives rise to an about $11\ MB$ event. The mask step is used to reduce the event size to the level of $250\ kB$ by masking the readout channels with specific patterns. The calibration time for entire HGTD is required to be less than five minutes. An estimation based on the required number of optical links shows that, the HGTD will need 48 FELIX I/O cards and 48 Data Handlers for the main data stream, and 32 FELIX I/O cards for the luminosity data stream. For the on-detector electronics, see other parts of the Section 4.4.

A detector control system (DCS) is being designed for the HGTD to ensure a coherent and safe operation. In general, it has two main tasks, monitor and control. In terms of the monitor, it can monitor the HGTD operational parameters such as the high and low voltage power supplies, the temperatures of the modules, peripheral electronics and cooling plate, the humidity and pressure of the hermetic volume, also detect any abnormal behaviors such as reaching critical conditions and detachment of hardware connections. In case of the abnormal behaviors, the DCS system should allow manual or automatic corrective actions. All the relevant monitoring parameters will be used for debugging, performance tuning and offline studies. In terms of the control, the DCS system can deliver control data stream through the high speed downlink to the entire on-detector electronics, thus set the detector to be in any desired operational state. The DCS elements are dispersively distributed all the way from the front-end electronics to the off-detector part. To implement the DCS system in the HGTD, the idea is to use the standard ATLAS DCS components everywhere as much as possible. For some special cases, dedicated strategy might be adopted. For example, due to the very limited physical space available, instead of using the standard DCS components, the DCS data stream share the same infrastructure with the on-detector electronics of the DAQ system, which is based on the lpGBT ASIC followed by the electro-optical module VTRx+ and the FELIX system. In addition, a dedicated finite state machine (FSM) for HGTD is being designed and will be integrated in the ATLAS FSM tree during data taking. The FSM can also be operated in stand-alone mode during the stage of commissioning and possible maintenance.

During the LHC period, most of the physics analyses are already suffering from the

systematic uncertainties on the measurement of the integrated luminosity delivered by the LHC, therefore a more precise measurement of the integrated luminosity is essential for achieving the physics goals of the HL-LHC. The mechanism of luminosity measurement is straightforward, which is to measure an indirect observable that is proportional to the instantaneous luminosity, or equivalently, to the pile-up number $\langle\mu\rangle$ (i.e. the average number of interactions per bunch crossing). The observables can be the average number of charged-particle tracks reconstructed in the inner tracker, the average number of the clusters in the pixel detector, the fraction of the bunch crossings without any interactions (event-counting method [149]), the average number of fired hits per bunch crossing (hit-counting method). A calibration technique called van der Meer (vdM) scans employed to determine the absolute delivered luminosity from the measured instantaneous luminosity of the bunch crossing reported the luminometer. The idea is to compute a scale factor by comparing the true luminosity value set by the accelerator operation group and the uncalibrated collision rate from the luminometer at the peak of the beam-separation scans. During the vdM scans, the pile-up number $\langle\mu\rangle$ of the beam is about 0.5 at the peak of the scans and 2×10^{-5} at the tail of the scans, so that there are only several tens of widely-separated bunched circulating in the LHC pipeline. During the HL-LHC operation, it's planned to use the same luminosity calibration method, which means that the luminometer should still be able to linearly response across seven orders of magnitude in $\langle\mu\rangle$, from the vdM regime to the high luminosity regime of $\langle\mu\rangle \approx 200$.

Chapter 5

Modular Peripheral Electronics Board Design

Prior to the PEB prototype, a PEB emulator system has been designed to help us verify some concerns in advance. The development of such a system is a crucial part of the whole PEB design chain. In addition, it can also greatly benefit other tasks of the HGTD, even become an important step on their own schedules.

5.1 Requirement analysis

As described in the Section 4.4.3, the design of the PEBs is extremely complicated and must have a comprehensive research and preparation before the initiation of the design. There will be several tens of ASICs and connectors mounted on each type of PEB, which has very limited areas available. With such a congested layout of components, the wire-routing of the printed circuit board (PCB) must be very challenging. A single lpGBT ASIC has up to 289 pins with the ball grid array (BGA) package and the pin pitch of just $0.5\ mm$. The mini-connector (FH26W flexible printed circuit (FPC) connector) for signal transmission and low voltage power supply has 71 pins with the pin pitch of $0.6\ mm$. The bPOL12V ASIC is integrated together with a piece of PCB as a DCDC converter with the footprint of $19\ mm \times 26\ mm$. On the PEB 1F, approximately, 11 lpGBTs, 55 71-pin FH26W connectors and 42 DCDC converters will be used, which certainly gives rise to a large electrical complexity of the PEB design. At the end, six types of PEBs with rather similar complexities have to be designed. As can be seen in the Section 4.4.3.2, the lpGBT ASIC is very complicated and with a lot of functionalities that are all critical to support the peripheral electronics. Making all the lpGBTs efficiently collaborate with each other and all the other components is already a demanding task, not to mention the unified design of the PEBs. As a pivotal part of the HGTD electronics, the PEBs interact with many other systems, such as the cooling plate, front-end electronics, DAQ and DCS system, which bring more complications to the PEB design.

Aside from the challenges coming from the complexity of the PEB, there are also some other concerns that may bring uncertainties and difficulties to the PEB research and development (R&D) and lead to possible delays on the agenda. The key ASIC lpGBT is completely a new chip not only to us but also to other users from CERN, therefore how it can be used on the PCB board, how it's configured to be capable of working in an optimal state and how it can work together with other ASICs need to be investigated, only a good understanding of the lpGBT is gained, can more confidence on a successful PEB design be achieved. Those key ASICs are still under the stage of development or characterization and heavily demanded by other LHC Phase-2 upgrade projects, so it is very likely that only a few ASICs or even nothing can be delivered to our hands. For example, at the beginning, we got no ASICs held in our hands even though they are urgently needed by this program, and it was expected that those key ASICs would gradually come to us over a long period of time with very insufficient quantities. Furthermore, the progresses of those ASICs are different, the MUX64 was a bit left behind and got delivered to us afterwards in the form of bare dies, which require many subsequent processes like packaging, irradiation test, temperature cycling test, etc, a few bPOL12Vs were also distributed to us later just in the form of packaged ASICs requiring further integration into DCDC converter modules. Another uncertainty came from the COVID-19 pandemic, which causes unexpected delays of the delivery business across the world and unpredictable lead time of products (PCBs, ASICs, electrical components, etc) from the manufacturers.

Due to the above-mentioned challenges facing the PEB design, the design of an emulator board (modular PEB) is proposed for the verification of the future PEB design even under the unfavorable situation at the stage of the PEB R&D. The modular PEB is expected to be a simplified version of the final PEB, resembling PEB as much as possible in many aspects such that as many concerns about the design as possible can be settled and as many related programs as possible can be assisted. With this intermediate board, it's foreseen that the PEB R&D will not be affected by those challenges and still moves forward with strong momentum. In order to fulfill the stated objectives, the following requirements have to be carefully taken into account for the design of the modular PEB:

- The data transmission of the versatile link (lpGBT + VTRx+) should be able to be exercised with the modular PEB. The data transmission of the HGTD from the on-detector side to the off-detector side is based on the versatile link, consisting of the two high-speed electrical uplink and downlink of the lpGBT, and the optical uplink and downlink fibers of the VTRx+. The performance of the versatile link is the key to the high-quality data communication and determines the success of the peripheral electronics. Therefore, the versatile link has to be integrated on the board, serving as the interface with the optical fibers of the DAQ system;
- It should reserve the interfaces with the front-end modules; The 71-pin and 13-pin

FH26W connectors can serve as the interfaces, which can not only support the modules as the signal sources of the system but also practice the procedure of attaching the HGTD modules to the PEBs. Considering the terribly-possible absence of the HGTD modules, the HGTD module emulators have to be designed to mimic the real modules both in dimension and in electrical functionality. So the interfaces should be designed to support both the module emulators and the real HGTD modules;

- Multiple lpGBT configuration methods should be supported and thus exercised since all of them are indispensable all the way from the PEB R&D to the PEB commissioning and maintenance; It would be highly profitable to have a better understanding of the lpGBT configuration with the help of this modular PEB, which will definitely make the future activities related to the PEB easier and more efficient. The lpGBT can be configured via either its I^2C slave port or its high-speed downlink with the two internal control (IC) bits and two external control (EC) bits embedded, where the IC bits are used for configuring the current lpGBT while the EC bits are used for the configuration of the slave lpGBT connected to the current lpGBT. In particular, for the configuration with the I^2C slave port, a dedicated lpGBT configuration board has to be designed based on the I^2C protocol so that it's guaranteed that the lpGBT can be configured properly even without a working high-speed downlink during the debugging stage. Instead of using the existing lpGBT programmer piGBT [150], another lpGBT programmer is newly designed to provide the electrical isolation, which can avoid possible damages to the scarce lpGBTs caused by the surging current of the host computer;
- The lpGBT cascade mode with one DAQ lpGBT as master and with one lumi lpGBT as slave should be implemented. The slave lpGBT is controlled by the master lpGBT either through its EC port or through its I^2C port, also the reference clock of the slave lpGBT comes from the master lpGBT. This kind of lpGBT combination is actually the basic structure of the lpGBT-based peripheral electronics, and its feasibility, performance and reliability is vital to the design of the PEB;
- All the three transmission rates supported by the lpGBT electrical ports (E-port) (320 $Mbit/s$, 640 $Mbit/s$ and 1.28 $Gbit/s$) should be easily set and tested. The serialized data streams with different transmission rates flow through the FLEX cables from the timing and luminosity data pins of the front-end ASICs to the channels of the lpGBT E-ports, then the lpGBT E-ports process the data streams at the set rate. Each E-port has a bandwidth of 1.28 $Gbit/s$ and four channels (E-links) for receiving serialized data streams. Depending on the set rate of the E-port, the number of available channels can be different. At the rate of 320 $Mbit/s$, all four channels are available; At the rate of 640 $Mbit/s$, just the channel 1 and 3 are available; At the rate of 1.28 $Gbit/s$, only the channel 1 is available; To test all three data rates of the E-ports

in a thorough way, a special connection strategy has to be used to carefully handle the confusing connectivities between the modules and lpGBT ports. The three data rates are targeting at the HGTD modules located in the three radial regions (inner, middle and outer regions) of the HGTD active area, and they will be extensively used in various combinations in the PEB design, therefore those combination cases of data rates are required to be considered in the modular PEB design;

- The key ASICs (lpGBT, DCDC converter, VTRx+ and MUX64) should be replaceable from the board surface in case of failure. Since each of the key ASICs is crucial to keep the board in a functioning state, any failure of the ASIC will make the entire board wasted, which means even those other still-working ASICs can not be reused. In that case, we would lost all the components mounted on the board surface, which makes the terrible situation of lacking ASICs even worse. It is unwise to adopt a strategy like that for the verification board. If the ASICs can be replaced easily from the board when not working, the only loss we have will be just that broken ASIC. To achieve such a replaceable design, the ASIC should be wrapped into a pluggable daughter board that interfaces with the carrier board via connectors;

- The modular PEB should be compatible with commercial ASICs in case of the key ASICs shortage; There are already some commercial ASICs in the market that can mimic the functions of the key ASICs. In the worst scenario that all the key ASICs are absent at the very early stage, those corresponding commercial ASICs can be used on the modular PEB to realize most of the functions, which makes the development still progress forward. Thanks to the replaceable design, those commercial ASICs can also be designed in the form of daughter board and effortlessly integrated into the carrier board via connectors. Hence, to carry out the compatible design, in addition to the daughter boards design of the key AISCs, the daughter boards design of the commercial ASICs should also be required.

- The technologies used in the PCB production and key mechanical parameters should be adopted on the modular PEB as much as possible, so that their feasibilities can be validated in advance. One of the main technologies is called plated over filled via (POFV) structure, which allows the use of vias right under the BGA pads and thus alleviate the burden caused by the high routing density of the PCB. The feasibility of the POFV structure can further reassure and help us to build more confidence on the future high-density PEB design. The distance between two neighboring FH26W connectors is of great important in the module installation. A proper distance can not only make the installation easier to operate but also give more room to other components on the PEB.

5.2 Design strategy

5.2.1 General hardware structure

In order to fulfill the above-mentioned requirements, the modular PEB is designed as a modularized board, where its main body serves as a carrier board while its components are enclosed in the shape of daughter boards and can be easily attached to the carrier board through connectors. It is a lightweight version of the PEB but with all the necessary functionalities. As the bridge connecting the front-end modules and DAQ system, the modular PEB is designed to interface with both of them, and their implementations are also considered together with the design of the modular PEB. Figure 5.1 shows the block diagram of the modular PEB system. As can be seen, modular PEB system is consisted of three parts, including the frontend, the carrier board in the middle and the backend.

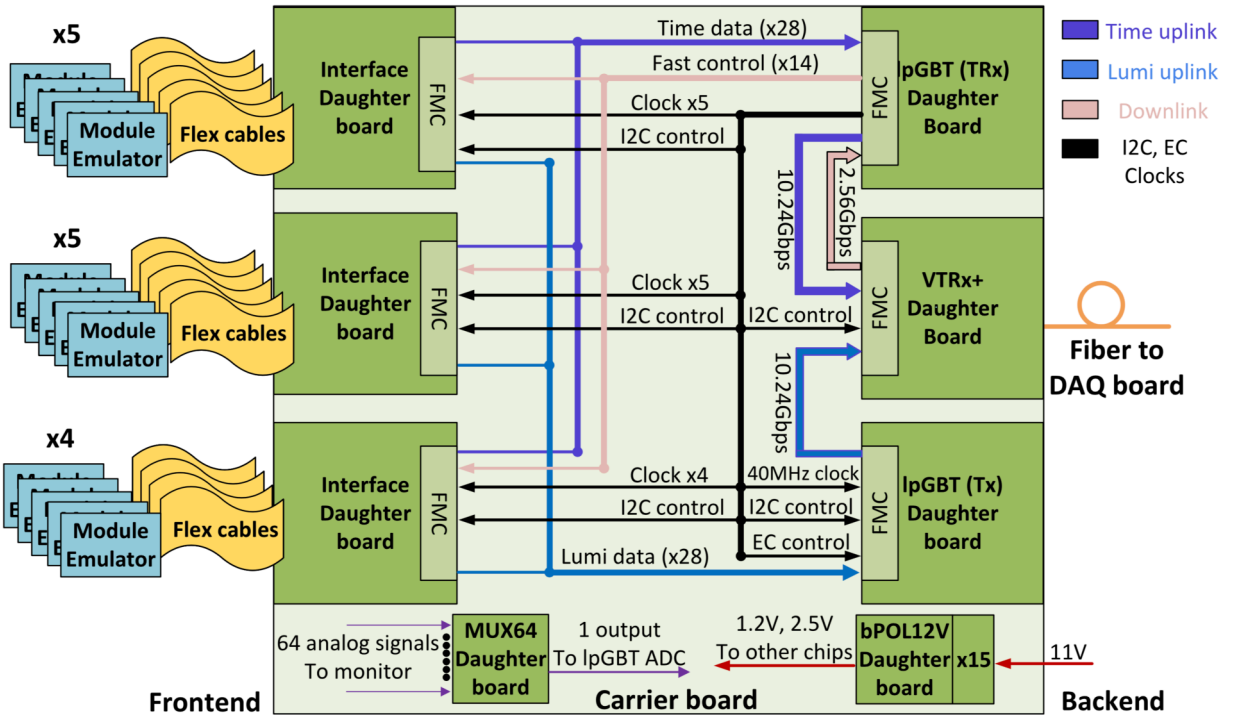


Figure 5.1: Block diagram of the modular PEB system [151]. Due to the limited number of the E-links of the lpGBT, the number of module emulators attached to the modular PEB is 14 rather than 15.

The carrier board part is actually our main dish, the modular PEB. Compared with the PEB, there will be a rather smaller quantity of components used on the modular PEB, including two lpGBTs, one VTRx+, one MUX64, 15 bPOL12Vs, 15 71-pin FH26W connectors and 15 13-pin FH26W connectors. As can be seen, all the components appear on the carrier board as daughter boards, and the ways they are connected with the carrier board are either through the FPGA Mezzanine Card (FMC) connectors or through the pin header connectors. For example, the three FH26W connector daughter boards (each mounted with

5 FH26W connectors) located on the left side of the carrier board, the two lpGBT daughter boards and one VTRx+ daughter board located on the right side of the carrier board are mounted with FMC connectors; The 15 bPOL12V daughter boards and one MUX64 daughter board located on the middle area of the carrier board are mounted with pin header connectors. Accordingly, on the carrier board side, there will be mounted corresponding connectors to match with those connectors on daughter boards. Those connector slots on the carrier board are not only for the daughter boards of key ASICs but also for the daughter boards of their commercial substitute ASICs. In this design, the substitute ASICs of the key ASICs lpGBT, VTRx+ and bPOL12V are respectively selected to be their commercial counterparts FPGA, SFP+ optical transceiver module (use SFP+ for simplicity) and TPS56428 to achieve the similar functionalities.

In the frontend part, there should be real HGTD modules or module emulators as the data sources of the system, and they are connected to the three FH26W daughter boards of the modular PEB through the FLEX cables. It's not expected that the real HGTD modules will be ready during the test of the modular PEB system, so the module emulators have to be designed to perform as the real modules. In terms of size, it is exactly as large as the real module, which is about $4\ cm \times 2\ cm$. In terms of function, it is built upon FPGA where corresponding logics can be implemented to be able to achieve what real modules can do. Also the dedicated FLEX cables designed by Johannes Gutenberg Universitaet Mainz are not expected to be in place during the joint test, so the commercial FPC cables matching the FH26W connectors will be used instead.

In the backend part, a DAQ board should be used to handle the data streams coming from the optical fibers of the modular PEB. The optical fibers are from the electro-optical converters, which can be either VTRx+ or SFP+ daughter board. Therefore, on the DAQ board side, to be able to receive the optical signals from the modular PEB, it should also support the electro-optical converters, which can be based on VTRx+, SFP+, or other types of devices. Since the requirement for the DAQ board is quite general, an existing MicroTCA fast control board is considered to be used instead of designing a new board.

5.2.2 Control and configuration strategy

Generally speaking, all the ASICs in the modular PEB system get controlled by configuring their registers, so the control and configuration are usually performed together. The control of the modular PEB system is built upon the DAQ lpGBT. The idea is that the DAQ lpGBT firstly gets configured either through its I^2C slave port or through its high-speed downlink, then broadcasts configuration messages to other devices (lumi lpGBT, VTRx+ and modules) either through its three I^2C master ports or through its specific E-ports. Figure 5.2 shows an overview of the control and configuration strategy of the system.

When received by the DAQ lpGBT, the serialized $2.56\ Gbit/s$ high-speed downlink DCS

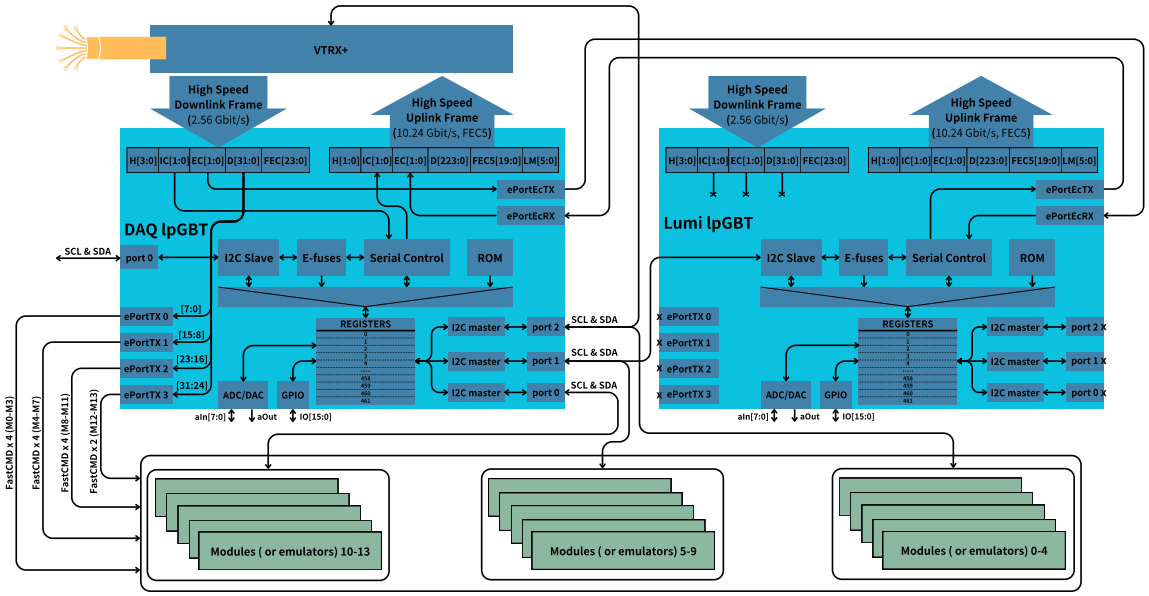


Figure 5.2: Block diagram of the control and configuration strategy of the modular PEB system.

data will be processed by the downlink logics and recover the configuration messages to be in the form of 64-bit data frames. The 64-bit data frame is comprised of 4-bit header for delimiting the start of the frame, two IC bits with a bandwidth of 80 *Mbit/s* for controlling the lpGBT itself, two EC bits also with a bandwidth of 80 *Mbit/s* for controlling the slave lpGBT connected with the current lpGBT, 32 data bits with a bandwidth of 1.28 *Gbit/s* carrying control information for the front-end modules, and the 24 forward error correction (FEC) for detecting and correcting transmission errors due to noise or single event upsets (SEU). Similarly, the uplink logics pack the serialized timing data from E-links into 256-bit data frames. The 256-bit data frame is also consisted of 2-bit header, two IC bits, two EC bits, 224 data bits, 20 FEC5 bit and six latency measurement (LM) bits for providing a rough estimation of the round-trip latency of the transceiver link (excluding E-links).

In terms of the fast control of front-end modules, the 32-bit configuration data embedded in the downlink frame of the DAQ lpGBT is split into four pieces (eight bits for each piece), which respectively go to the four downlink E-ports. Each E-port has the bandwidth of 320 *Mbit/s*, which precisely matches the eight bits also with the bandwidth of 320 *Mbit/s*. Given each E-port has four E-links, all four E-links are available when the data rate is set to be 80 *Mbit/s*, just the first and the third E-links are available when it's 160 *Mbit/s*, and only the first E-link is available when it's 320 *Mbit/s*. A mirror function is provided for the downlink E-ports. Once activated, in the same E-port, the second, third and fourth E-links repeat the data of the first E-link in the case of 320 *Mbit/s*; Also in the case of 160 *Mbit/s*, the fourth E-link repeats the data of the third E-link while the second E-link repeats the data

of the first E-link. In the case of 80 *Mbit/s*, the mirror function is not supported because all four E-links are already occupied. In the scenario of modular PEB design, the mirror function will be activated and the data rate will be set to be 320 *Mbit/s*, therefore, there will be 16 320 *Mbit/s* downlink E-links, the first 14 of them are connected with the 14 front-end modules for delivering the fast commands. Because of using the mirror function, the modules connected to the same E-port will receive exactly the same fast command. The fast command is embedded into the 320 *Mbit/s* serialized data stream, and it will be recovered and interpreted as actual operations by the electronics of the modules. Since the lumi lpGBT will be working in the simplex transmitter mode, the high-speed downlink is closed and the fast control is not supported by this lpGBT, thus its downlink E-ports will be all disabled.

In terms of the slow control of the modular PEB system, there are generally three control paths all starting from the DAQ lpGBT. The first one is the IC path. The two downlink IC bits with the bandwidth of 80 *Mbit/s* flow into the serial control logic of the DAQ lpGBT, which extracts the configuration messages and implement corresponding operations either to the E-fuses or to the registers. For the E-fuses, they can be permanently programmed and will be loaded into the registers in the future powering up. For the registers, their values firstly get modified, then related signals are propagated to other parts of the ASIC, thus actions are taken and ASIC status is changed. Actually, after the change of the registers, the three I^2C master blocks can also be affected since they are controlled by some specific registers. The three blocks correspond to three ports, which are used to control other devices. For example, the modules 10-13 share the bus lines of the port 0; The modules 5-9 and the lumi lpGBT share the bus lines of the port 1; The modules 0-4 and the VTRx+ share the bus lines of the port 2. There are also corresponding IC feedback information going backwards and being wrapped as two bits into the high-speed uplink data frame of the DAQ lpGBT. The second one is the EC path. The two downlink EC bits with the bandwidth of 80 *Mbit/s* successively goes through the EC ports of the DAQ and lumi lpGBTs, and finally flow into the serial control logic of the lumi lpGBT as well. The procedure after this is exactly the same with that of the first control path. The only difference is that the three master ports of the lumi lpGBTs are not used at all. The third one is the I^2C path. An external I^2C master device, which can be the lpGBT programmers like UPL and piGBT, is connected to the I^2C slave port of the DAQ lpGBT and communicate with the related slave logic. Like the serial control logic, the slave logic can also be able to handle both the E-fuses and registers in the same way, of course, the three master ports can also be reached and well controlled by the external master device.

In addition, the enable signal of the 12 bPOL12Vs serving for the modules are controlled by the general purpose input/output (GPIOs) of the DAQ lpGBT. There are 16 GPIOs in total, but just five of them (from 7th to 11th pins) are used for this purpose, which means several bPOL12Vs share one enable signal coming from a single GPIO. The GPIO

logic is controlled by the registers, which can be configured by the three above-mentioned slow control paths. Another five of them (from 12th to 16th pins) are connected to the 14 modules for generating their global reset signals. Similarly, several modules will also share one reset signal because of the limited number of GPIOs.

5.2.3 Connectivity of timing and luminosity data streams

The connectivity of timing and luminosity data streams is certainly one of the most complex parts of the modular PEB design. Each module has two timing data streams and two luminosity data streams. In this design, all the luminosity data streams are routed to the lpGBT no matter what timing data rate the module will be possibly working with. The timing data streams are routed to the DAQ lpGBT while the luminosity data streams are routed to the lumi lpGBT. It's required that all the serialized data streams of the 14 modules should be strategically routed to the E-ports of the two lpGBTs such that we can fully exploit the test ability of this design to validate the data transmission related concerns as much as possible, such as the tests of the data rate settings and different connectivity patterns. Figure 5.3 shows the connectivity strategy of timing data streams. Since the connectivity of the luminosity data streams is following the exactly same strategy with the that of the timing data streams, it will be not be repeated.

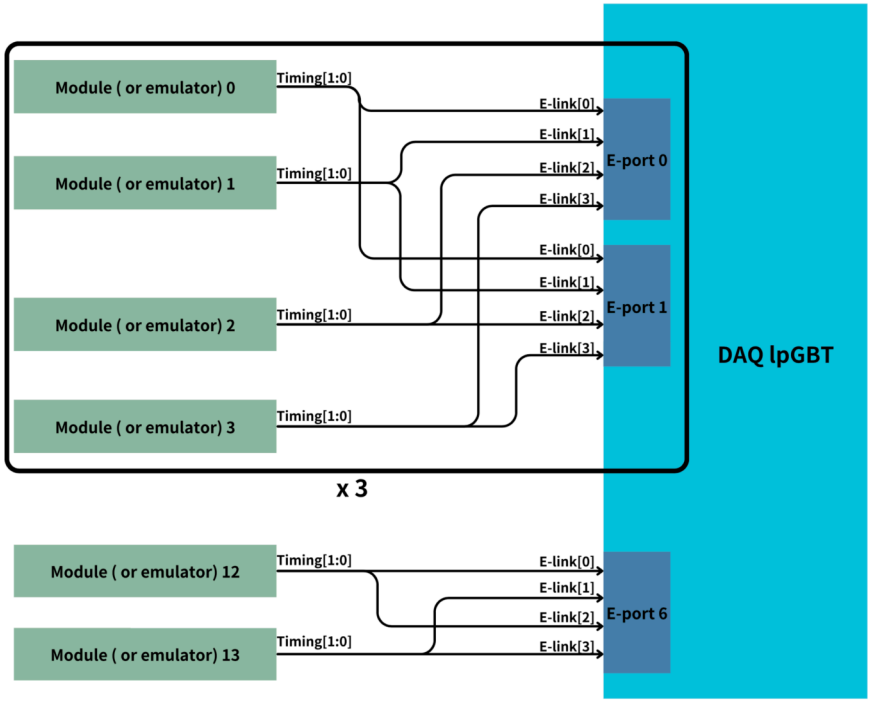


Figure 5.3: Connectivity strategy of timing data streams.

Each lpGBT has seven configurable uplink E-ports, and each E-port has four E-links. In total, there are up to 28 E-links, which can exactly match the 28 timing data streams of the 14 modules. Like the downlink E-ports, each uplink E-port also has a fixed bandwidth

(1.28 $Gbit/s$) and priority level across its four E-links, which should be carefully taken into account when doing the connectivity. The E-port can support three data rates, 320 $Mbit/s$, 640 $Mbit/s$ and 1.28 $Gbit/s$, once a certain data rate is set, all the E-links of this E-port can only work in that data rate. Mixed data rates within one E-port are not supported, but they can be used across different E-ports. All the four E-links are available when the E-port is set to be 320 $Mbit/s$, just the first and third E-links are available when the E-port is set to be 640 $Mbit/s$, only the first E-link is available when the E-port is set to be 1.28 $Gbit/s$.

Specifically speaking, the first 12 modules are averagely divided into three groups, and each group is handled by two E-ports in the same way. As shown in Figure 5.3, in one group, the two timing data streams of each module go to two E-links belonging to different E-ports, also the two E-links have the same priority level (or ranking) in their own E-ports. Therefore, the two E-ports should be set at the same data rate, thus the four modules should also be working at the same data rate. This kind of connection can guarantee that, no matter what data rate the two E-ports is set to be, their bandwidths (1.28 $Gbit/s \times 2$) can be always fully used thus they can be able to support the maximum number of modules. For example, when working in the data rate of 1.28 $Gbit/s$, they can support one module, which actually already uses up all the bandwidths of them; When working in the data rate of 320 $Mbit/s$ or 640 $Mbit/s$, they can support all the four modules in this group. The 12 modules already take 6 E-ports of the lpGBT, Therefore, the last two modules (the thirteenth and fourteenth) have to be connected with the 7th E-port. The timing data streams of the 13th module are attached to the 1st and 3rd E-links while the timing data streams of the 14th module are attached to the 2nd and 4th E-links. Similarly, this kind of connection can also make sure that, in whatever data rate the E-port is running at, the bandwidth of the 7th E-port is completely used. For example, in case of 640 $Mbit/s$, the E-port can support the 13th module, while in case of 320 $Mbit/s$, the E-port can support both of the modules.

With the connectivity strategy, plenty of data rate patterns for the data streams can be performed on the modular PEB. Table 5.1 lists some typical patterns, which are applicable for both the the timing and luminosity data streams.

5.2.4 Power supply

The modular PEB system should be supplied with low voltages (LV) and high voltages (HV). The former is for supplying all the electronics of the system while the latter is for providing bias voltages of the LGAD sensors. The HV power supply design is straightforward. It will firstly get the working voltages from external HV power supply crates then directly feed them to the front-end LGAD sensors without any conversion. The LV power supply design is comparatively complex. After getting the external LV input (10 V), the DCDC converters, either bPOL12V or its commercial substitute TPS56428, will convert it to the targeting voltages, including 1.2 V, 2.5 V and 3.3 V, which will be further distributed to all the

Table 5.1: Typical data rate patterns for the timing or luminosity data streams. The values are the data rates of the modules; "....." means the module is not supported in that pattern.

Module Number	Pattern I	Pattern II	Pattern III	Pattern IV
M0	320 Mbit/s	640 Mbit/s	1.28 Gbit/s	1.28 Gbit/s
M1	320 Mbit/s	.....	.....	.....
M2	320 Mbit/s	640 Mbit/s	.....	.....
M3	320 Mbit/s	.....	.....	.....
M4	320 Mbit/s	640 Mbit/s	1.28 Gbit/s	640 Mbit/s
M5	320 Mbit/s	.....	.....	.....
M6	320 Mbit/s	640 Mbit/s	.....	640 Mbit/s
M7	320 Mbit/s	.....	.....	.....
M8	320 Mbit/s	640 Mbit/s	1.28 Gbit/s	320 Mbit/s
M9	320 Mbit/s	.....	.....	320 Mbit/s
M10	320 Mbit/s	640 Mbit/s	.....	320 Mbit/s
M11	320 Mbit/s	.....	.....	320 Mbit/s
M12	320 Mbit/s	640 Mbit/s	320 Mbit/s	640 Mbit/s
M13	320 Mbit/s	.....	320 Mbit/s	.....

other ASICs in a certain way. As described in Section 4.4.3.3, the bPOL12V is a step-down buck DCDC converter, which has the advantages of high efficiency and low power dissipation compared with the other kind of voltage conversion technique, low-dropout regulator (LDO). Therefore, its substitute ASIC TPS56428 is also selected to be a step-down DCDC converter with a good performance. Figure 5.4 shows the block diagram of the power supply design for the modular PEB system.

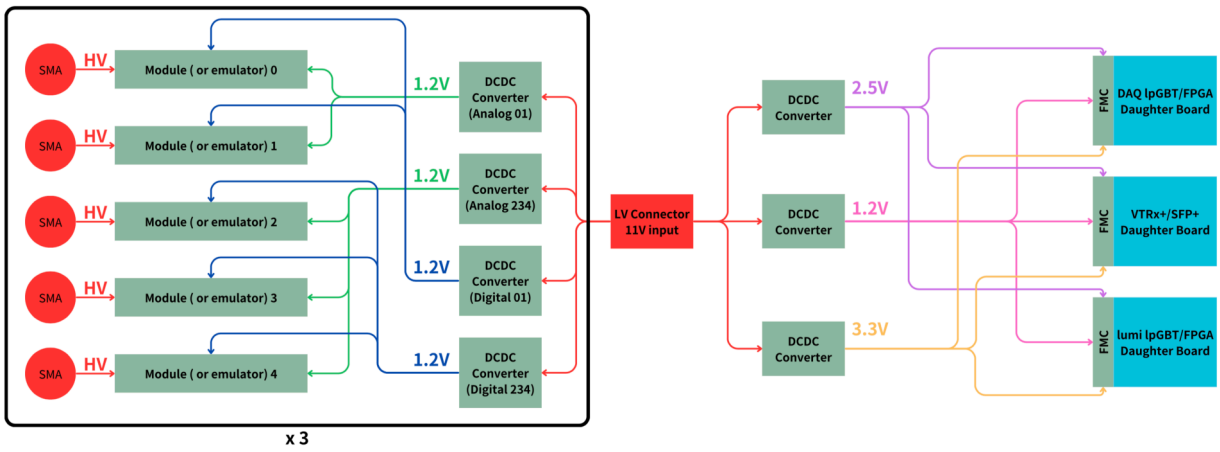


Figure 5.4: Block diagram of the power supply design for the modular PEB system.

A single LV power Molex connector called 39-30-1040, followed by an electro-magnetic interference (EMI) suppression filter, is mounted on the carrier board receiving the input voltage of 11 V. The idea is to firstly convert the input voltage to different voltages with DCDC converters then distribute the converted voltages to the connectors of daughter boards. The input voltage is shared by all the 15 DCDC converters. Three of them are used for serving

the ASICs on modular PEB. Their converted voltages are respectively 1.2 V, 2.5 V and 3.3 V, which further go to the one pin header connector and three FMC connectors respectively for the MUX64, two lpGBTs and VTRx+ daughter boards for their power supplies. The remaining 12 of them are used for serving the HGTD modules (or emulators). Their converted voltages are just 1.2 V, which all go to the three FMC connectors for the FH26W daughter boards. Each of the FMC connectors is connected with four DCDC converters for supplying five modules that are attached to this FH26W daughter board, two DCDC converters are for the analog power supply of the five modules while the other two are for the digital power supply of them. Specifically, for the analog power supply, the first two modules share one DCDC converter while the other three modules share another one DCDC converter, the same sharing strategies are also used for the digital power supply. Hence, such two-module and three-module sharing strategies can be evaluated on the modular PEB board, and the performance of them will help us to determine how the power supply is carried out for the PEB board.

Each module requires one HV channel, so 15 HV channels have to be provided for the 15 modules that are attached to the three FH26W daughter boards. To do that, each FH26W daughter board is mounted with five SubMiniature version A (SMA) connectors for the HV power supply input. Then each HV input will be routed to its corresponding 13-pin FH26W connector on the FH26W daughter board. Finally, the HV will be further delivered to the same 13-pin FH26W connector on the module through the FLEX cable.

It's worth mentioning that the power supply strategy described here is just a general overview for the whole modular PEB. How the power supply of each daughter board and module emulator is handled will be described in Section 5.3.

5.2.5 Monitoring

The monitoring part of the modular PEB system is build on the multiplexer MUX64, the ADCs and GPIOs of the two lpGBTs. The idea is that all the signals to be monitored in this system are firstly directly or indirectly routed to the ADCs and GPIOs of the two lpGBTs, then the values of the corresponding registers are refreshed accordingly and read out by the DCS system. shows an overview of the monitoring strategy.

The signals being monitored include the five voltages (related to analog current return voltage, delivered analog and digital voltages, two voltages from the ALTIROC temperature sensors) of each module, the temperature and received signal strength indicator (RSSI) of the VTRx+, the power good signals of all the 15 DCDC converters, two temperatures on the carrier board, the input LV to the system, the voltage supply for the ASICs on modular PEB, the temperature of the first module, and the temperatures of the first two DCDC converters. Because the MUX64 input pins are not enough, one of the 13th module temperatures is not able to be monitored, also the five voltages of the 14th module are

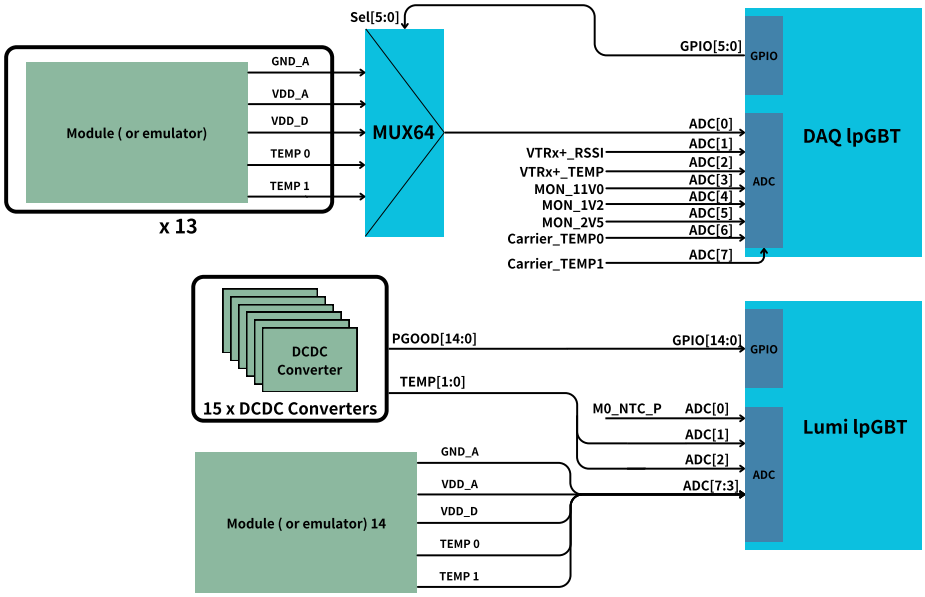


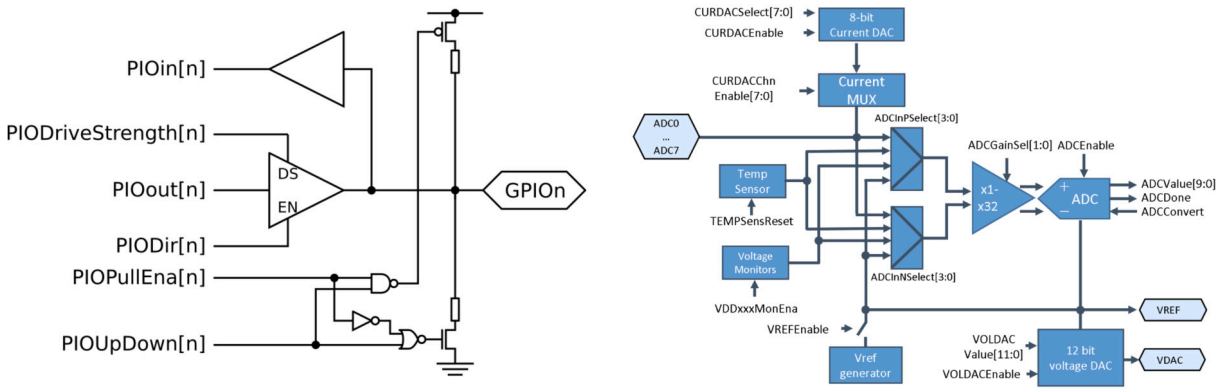
Figure 5.5: Block diagram of the monitoring strategy.

routed to the last ADC pins of the lumi lpGBT. The six selection pins of the MUX64 are connected with the first six GPIO pins of the DAQ lpGBT, so that the strobe signal to input channel can be controlled by DCS system. There are two kinds of temperature monitorings in the system. One is monitoring the ASIC internal temperature, which can be done by directly reading out the voltages of temperature sensors embedded in the ASICs, such as the temperature monitoring of the ALTIROCs, bPOL12Vs and VTRx+. The other one is monitoring the board temperature, which can be carried out based on external thermistor sensors mounted on the board surface, such as the temperature monitoring of the carrier board and module flex. The chosen thermistor sensor is a surface mounting technology (SMT) type component with high accuracy and large operating temperature range ($-40^{\circ}C$ to $120^{\circ}C$). It is used together with a resistor in series, and the $1.2\ V$ voltage is applied on them. Finally, the voltage drop on the thermistor sensor will be fed to the lpGBT ADC.

The ADC and GPIO are the critical parts of the monitoring chain. The GPIO is used for generating or receiving digital pulse, such as generating selection pulses for MUX64 and sensing the power status of DCDC converters, and it is also controlled by the lpGBT registers. Figure 5.6(a) shows the block diagram of one GPIO structure. As can be seen, it is high configurable. Each GPIO pin can be set as input or output with configurable CMOS driver and pull settings. In addition, all pins operations are synchronous with the internal system clock ($40\ MHz$).

The ADC is used for converting all the monitored voltages into digital values, so it determines the performance of the entire monitoring system. Figure 5.6(b) shows the block diagram of the ADC logic. It is a differential 10-bit successive approximation register (SAR)

ADC with the input dynamic range of from -VREF to VREF, where VREF can be either generated by the built-in reference voltage generator or externally provided by the VREF pin. More specifically, the logics that need reference voltage are the two input MUXes, ADC converter and the additional DAC converter. The inputs of the ADC can be divided into two types, external inputs and internal inputs. The eight ADC pins belong to the former while the built-in temperature sensor and voltage monitors belong to the latter. All channels, including all the inputs and the reference voltage, are connected to the positive and negative input MUXes. By configuring the two related registers respectively for the two input MUXes, both the ADC positive input and negative input can be selected from the channels. Then they are amplified by the amplifier with configurable gain settings of 2x, 8x, 16x, 32x, after which the voltage difference between them is converted to binary value the ADC logic. In addition, the ADC pins (ADC[7:0]) feature a configurable current source, whose output current intensity can be programmed by the 8-bit current DAC. The output current can also be attached to any of the ADC pins by setting a dedicated register. This feature can be used to measure external resistance connected to the ADC pin, given the fact that the voltage drop on the resistor is provided by the ADC and the current flowing through the resistor is also known.



(a) GPIO (b) ADC

Figure 5.6: Block diagram of the lpGBT GPIO (a) and ADC (b) [145].

5.2.6 Clock network

All the clocks used in the modular PEB system are originated from the DAQ lpGBT. The primary clock of the lpGBT can either be recovered from the 2.56 *Gbit/s* downlink high-speed data stream or directly come from an external 40 *MHz* reference clock. After it's further processed by the clock generator, clocks with the frequency of from 5.12 *GHz* down to 40 *MHz* are generated for the use of the full chip. Figure 5.7(a) shows the clock generator of lpGBT. As can be seen, the clock generator consists of a phase locked loop (PLL), a 1/2 clock divider and a 1/64 clock divider, and the PLL is comprised of phase detector (PD),

frequency detector (FD), phase and frequency detector (PFD), low-pass filter (LF) and LC-tank voltage controlled oscillator (VCO). When recovering clock from data, the PD and FD blocks are used and the $2.56\ GHz$ clock right after the 1/2 clock divider will be fed back to the PD and FD as in-phase quadrature (I/Q) signal. When using $40\ MHz$ reference clock, the PFD is used and the PLL acts as a frequency multiplier, and the $40\ MHz$ clock after the 1/64 clock divider will be fed back to PFD. Finally, the lpGBT can output up to 29 electrical clocks (E-clock), whose frequency, drive strength and pre-emphasis are configurable.

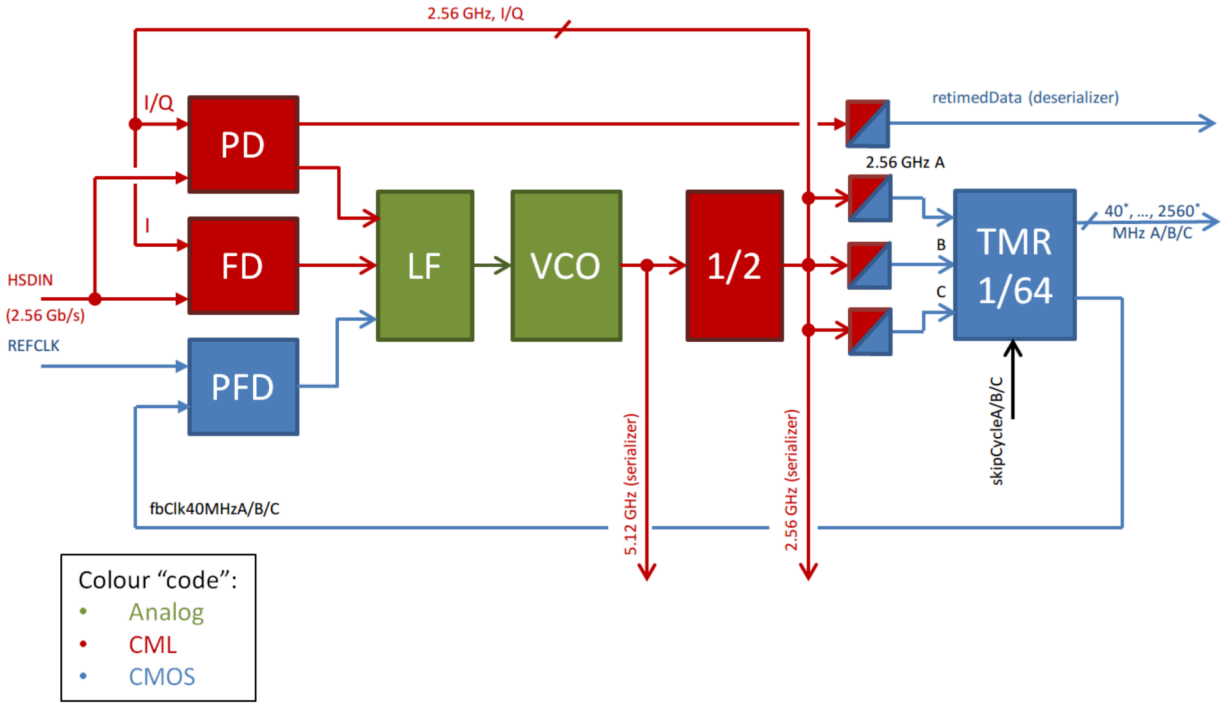
In addition to the clock generator, there is also a phase shifter circuit, Figure 5.7(b) shows its block diagram. Firstly a 7-to-1 MUX can strobe one of the seven input clocks, including one logical zero voltage and six different clocks of $40, 80, 160, 320, 640, and 1280\ MHz$, and the strobing pulse can be configured by users. After the MUX, two delay units are used in series, the first one is with a coarse delay cell of $800\ ps$ while another one is with a fine delay cell of $50\ ps$, which can be configured to be enabled or not. The number of delay cells activated in both of the delay units can also be configured. Finally, the lpGBT can also provide four phase shifter clocks, which are programmable in phase, frequency, drive strength and pre-emphasis.

Figure 5.8 shows the clock network of the modular PEB system. The E-clocks are used to provide clocks for the front-end modules (or emulators), each of which just requires one E-clock with the frequency of $40\ MHz$ or $320\ MHz$. They are delivered all the way from the DAQ lpGBT on the modular PEB to the frontend successively through the wires on PCB, FH26W connectors on daughter boards, FLEX cables and again FH26W connectors on modules. The PEB will also adopt such kind of clock path, so it will be largely beneficial to the PEB design if the clock path can be evaluated with the modular PEB. The parameters of those E-clocks can be adjusted accordingly during the joint test based on the actual demands of the front-end electronics. One of the phase shifter clocks of the DAQ lpGBT will be programmed to be $40\ MHz$ as the external reference clock of the lumi lpGBT, whose E-clocks and phase shifter outputs will not be enabled.

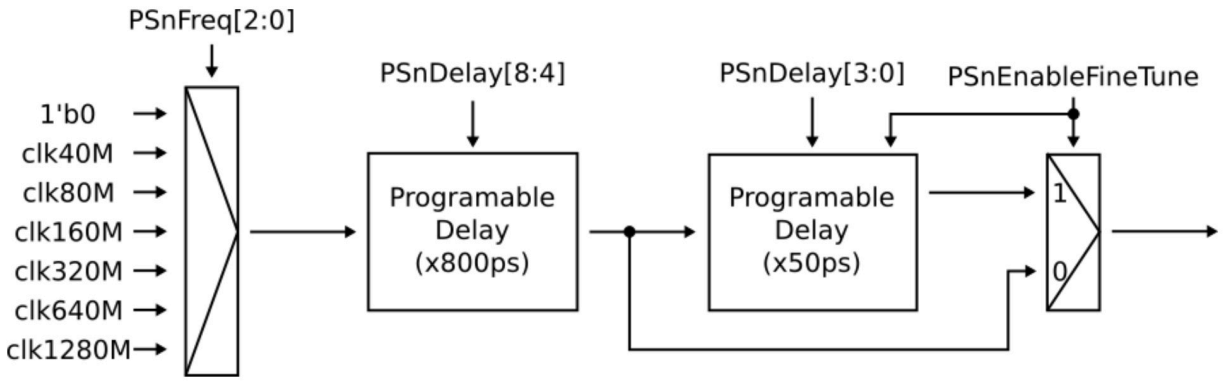
5.2.7 Other considerations

The above-mentioned design strategies are basically described based on the key ASICs. In the case of using substitute ASICs, either fully or partially, the strategies are still applicable. In the aspect of hardware connection, both the key ASIC and its substitute are integrated in daughter boards and adopt the common FMC connectors, therefore, now matter what kind of chip packagings the substitutes use, to the carrier board, it's the same. In the aspect of functionality, the substitutes are carefully chosen such that they have the similar electrical functions and performance or can mimic the functions of the key ASICs.

The substitute ASIC for lpGBT is a Xilinx KINTEX-7 FPGA. It has about 480,000 logic cells and up to 32 high-speed transceivers with the bandwidth of $12.5\ Gbit/s$, which



(a) Clock generator



(b) Phase shifter

Figure 5.7: Block diagram of the clock generator (a) and phase shifter (b) [145].

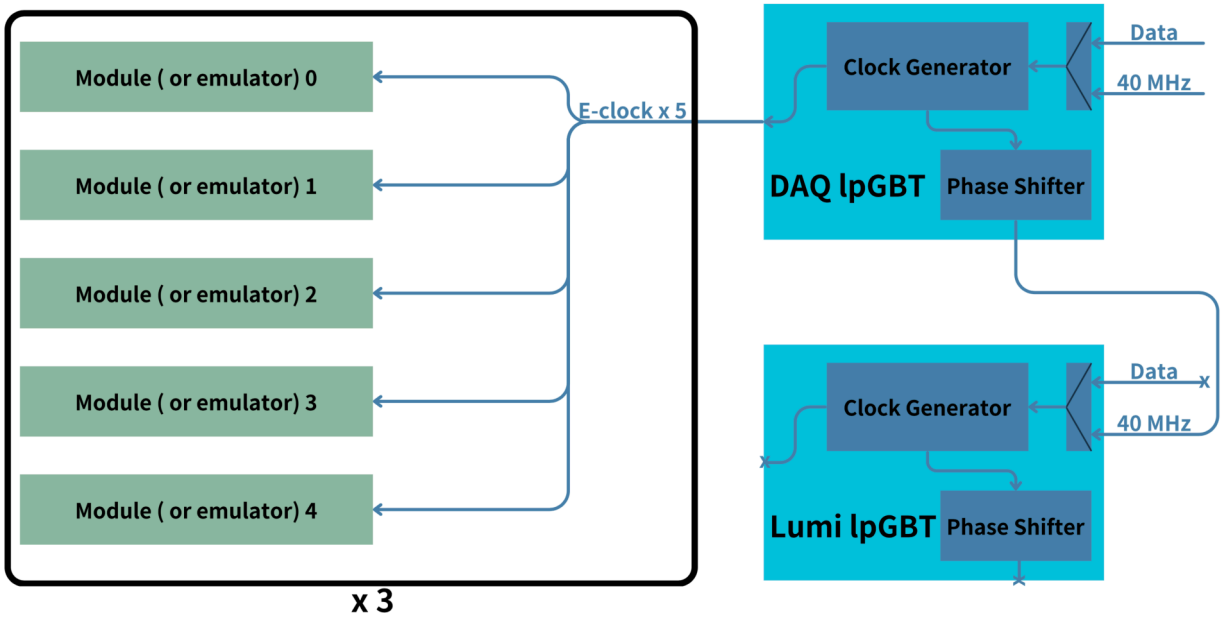


Figure 5.8: Block diagram of the clock network of the modular PEB system.

can easily fulfill the requirements on the lpGBT logics implementation and the data rates of high-speed uplink and downlink. The substitute component for VTRx+ is the SFP+ optical transceiver module, which can do the eletro-optical conversion at the data rate of up to 10.5 $Gbit/s$.

As described in Section 5.2.4, the substitute ASIC for bPOL12V is the commercial DCDC converter TPS56428, which can provide the target voltages by adjusting those resistors mounted on the daughter board. The TPS56428 daughter board is described in Appendix B.1. As mentioned in Section 5.2.1, the substitute of the real HGTD module is a dedicated module emulator, which should be designed to behave as the module before they are well produced and delivered to us. Section 5.3.3.1 presents the module emulator in details.

5.3 Hardware design

5.3.1 Carrier board

The carrier board is the pivot of the modular PEB since it handles the connectivities between all those components. The wire routing of this board can be taken as an implementation of the design strategy, and it is not only high-density but also with very large complexity. Figure 5.9 is a top view of the carrier board. For a more clear view of the carrier board, the picture was taken when most of the connector were not mounted.

The carrier board is a rather large rectangle board with the size of $23\ cm \times 12.2\ cm$. There is no any ASIC mounted on the board and just the slots for different types of connectors. Those slots include the right three for the FMC connectors of the FH26W daughter boards,

the left three respectively for the FMC connectors of the DAQ lpGBT, VTRx+ and lumi lpGBT daughter boards from top to bottom, the central one for pin header connector of MUX64 daughter board, the fifteen in the middle region for the pin header connectors of bPOL12V daughter boards, and other minor slots for the connectors of testing, debugging and LV power supply. In addition, there will be six green LEDs indicating the attachments to the six FMC connectors. Those fifteen shiny pads around the bPOL12V daughter boards slots are actually the exposed copper layer of the PCB, and they are not only serving as the ground of the board but also helping the heat dissipation of the DCDC converters. There are two drilling holes on each side of a FMC connector, which are used to mechanically fasten the daughter board to the carrier board. In addition, another four mechanical holes are drilled on the corners for installing support pillars to the board.

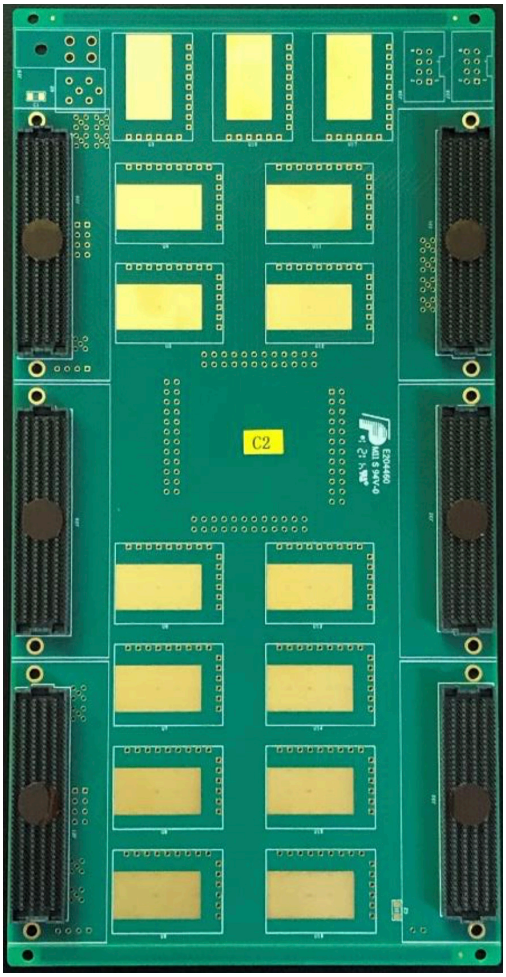


Figure 5.9: Top view of the carrier board.

The FMC connector plays an important role on the carrier board. The FMC is actually a VITA (VMEbus International Trade Association) 57.1 standard [152] that defines the daughter board (I/O mezzanine module) with connection to an FPGA or other device. This standard specifies the so-called FMC connector, which has a low profile and high pin density. It has ten rows (labelled as A, B, C, D, E, F, G, H, J, K) with 40 pins each, and pins are

specified for handling different signals, such as gigabit data and clocks, user data and clocks, I^2C, JTAG, power supply, and etc. In terms of the pin assignment of the FMC connectors, the carrier board design as well as the related daughter boards design follows the standard as much as possible. In addition, the standard also includes a methodology where FMC mezzanine modules (i.e. daughter boards) must provide its hardware information that can be read by an external controller (like FPGA for most of the time) on the FMC carrier platform (i.e. carrier board). The hardware information should be stored in a memory chip of the FMC mezzanine module in a specific format defined by the FRU (field replaceable unit) information storage definition, as described in [153].

In the carrier board design, there are two most challenging parts. One is the connectivity strategy of all the components. It has to be carefully considered that how the connections in different aspects are handled, and those aspects include power supply, clock, control, monitoring, timing and luminosity data, which are already discussed in Section 5.2. The other one is the high-density wire routing. In order to carry out the connectivity strategy, the board is designed with 14-layer stack-up, 6 layers for wire routing, 4 layers for power and 4 layers for ground. There are more than 1800 connections and 32 power shapes, which mean the large complexity and heavy workload of the board design. From this point of view, the design can be a good practice of the PEB design, which will adopt the similar design strategy and might be several times replications of this design.

5.3.2 Daughter boards

5.3.2.1 lpGBT daughter board

The lpGBT daughter board holds the lpGBT and serves as a FMC mezzanine module for convenient integration of lpGBT to the carrier board. The design idea is that all the lpGBT pins are routed to the FMC connector such that the carrier board can have full communication with the lpGBT once the lpGBT daughter board is plugged.During the design of this board, it's not guaranteed that which version of the lpGBT (v0 or v1) might be mounted on the board at the end, so the board is designed to be compatible with both v0 and v1 thanks to the unchanged pinout structure of the lpGBT. Considering its individual test before integration to the carrier board, it is also designed to be compatible with the Xilinx evaluation board so that it can be readily tested right after its production. The design can support the lpGBT to work as both DAQ and lumi modes, therefore just one type of the lpGBT daughter board is needed.

Figure 5.10 shows the top and bottom views of the board. On the bottom side, the lpGBT is mounted right in the middle while the FMC connector is on the side. On the top side, there are other minor components, such as the reset button, the dip switches for the settings of lpGBT address, working mode, boot flow and PLL lock mechanism, 14-

pin through-hole header connector for lpGBT configuration via the dedicated programmer UPL, and an electrically erasable programmable read-only memory (EEPROM) for storing the ”inventory” information of this board. The header connector is selected to be the same type with the CERN embedded monitoring and control interface (EMCI) board, and the pin assignment is also the same, so the facilities developed for EMCI board can be used by the lpGBT daughter board as well.

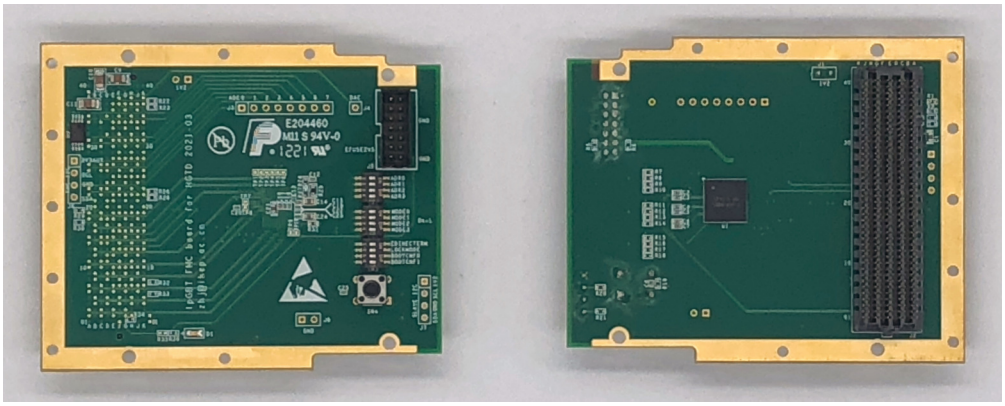


Figure 5.10: Top (left) and bottom (right) views of the lpGBT daughter board.

As described in Section 4.4.3.2, the lpGBT is the most complicated ASIC in the HGTD electronics, and it has up to 289 (17×17) pins with the ball grid array (BGA) package and the pin pitch of just $0.5\ mm$, which poses great challenge on the wire routing of the PEB design that may hold up to 12 lpGBTs. As discussed in Section 5.1, the PEB will adopt the POFV structure, which allows application of vias right under the lpGBT BGA pads. Therefore, this lpGBT daughter board is designed with this technology for validating its feasibility even though the space on the board is quite adequate. By using POFV on lpGBT pads, the fanout becomes more flexible. With traditional way, all the 289 pads have to be firstly fanned out to a nearby location on the surface then use vias to connect with other layers if needed. But with the POFV, the pads can be directly connected with any layer through the vias right under themselves, which avoids the annoying fanout of the pads thus simplify the wire routing.

The way the lpGBT pins are routed to the FMC connector pins should adopt the VITA 57 standard as much as possible. Especially the lpGBT pins of power supply, high-speed data transmission, I^2C, phase shifter clocks and reference clock should be connected to the corresponding pins of the FMC connector, since the board should be compatible with Xilinx evaluation board, whose FMC connector also adopts the same standard. For the other pins, such as pins of E-links (uplink, downlink and clock), EC-ports, GPIO, ADC and configuration are connected with those connector pins dedicated for user data in a more flexible way, for example, the connections can be adjusted anytime during the design for easier wire routing. In order to solve the possible I/O standard mismatch between the lpGBT and FPGA pins, $0.01\ uF$ AC coupling capacitors are used in series with the pins

of reference clock, high-speed uplink and downlink data transmissions. The READY pin is cascaded with a 220 Ω resistor and a green LED to the ground, so that it's clear when lpGBT is in the ready state. The 2.5 V power supply pins for E-fuses programming are just connected to the header connector. The pin assignment of this connector is presented in details in Section 5.3.3.2. The reset pin is connected to not only an external manual reset circuit with a button but also the header connector. On the board, there is also a thermistor sensor whose voltage output is connected to the last ADC pin, which is already taken by the voltage output of the thermistor sensor on carrier board, therefore, a 0 Ω resistor is used in series in between for helping select which temperature to monitor.

There are another twelve pins connected to three dip switches for setting some general parameters in a hard-wired way. Figure 5.11 shows a close-up of the dip switches. Among them, four address pins are used to set the four least significant bits (LSBs) of lpGBT chip address used in I^2C or serial control (IC and EC) configuration while the three most significant bits (MSBs) are set by the register called ChipAddressBar. Four mode pins are used to set the working mode of lpGBT, and each of the sixteen combinations correspond to a working mode characterized by the data rate and encoding method of the high-speed uplink data transmission, and the transmit-receive type. Table 5.2 lists all the sixteen working modes and their corresponding mode pins settings. The remaining four pins include the LOCKMODE pin for determining PLL lock mechanism (40 MHz reference clock locking or reference-less locking), the EDINECTERM pin for enabling the 100 Ω termination resistor of the EC-port receiver, the BOOTCNF0 and BOOTCNF1 pins for determining where and how the register values are loaded at the beginning of the lpGBT booting-up. In addition to the EDINECTERM pin, the termination of the EC-port receiver can also be activated by the ASIC register configuration. It's worth mentioning that, on the modular PEB, the DAQ lpGBT should work as a master at the MODE[3:0]=4'b1011 (transceiver, 10.24 $Gbit/s$ and FEC5) while the lumi lpGBT should work as a slave at the MODE[3:0]=4'b1001 (simplex transmitter, 10.24 $Gbit/s$ and FEC5).

5.3.2.2 VTRx+ daughter board

The VTRx+ daughter board is designed to hold the VTRx+ module and route its pins to the FMC connector. The design should also consider the scenario that the board is tested by a Xilinx evaluation board before it gets integrated into the carrier board. Figure 5.12(a) shows a picture of this daughter board. As can be seen, the connector on the left side is the specified connector for holding the VTRx+ module and it has 40 pins with the pitch of 0.4 mm and exactly the same pin assignment with the VTRx+. The FMC connector is located on the right side. In the area close to the FMC connector, there are slots for pin header connectors, LDO and level translation ASICs. The two metallic pads beside the VTRx+ module connector are used for installing the heat dissipation structure of the

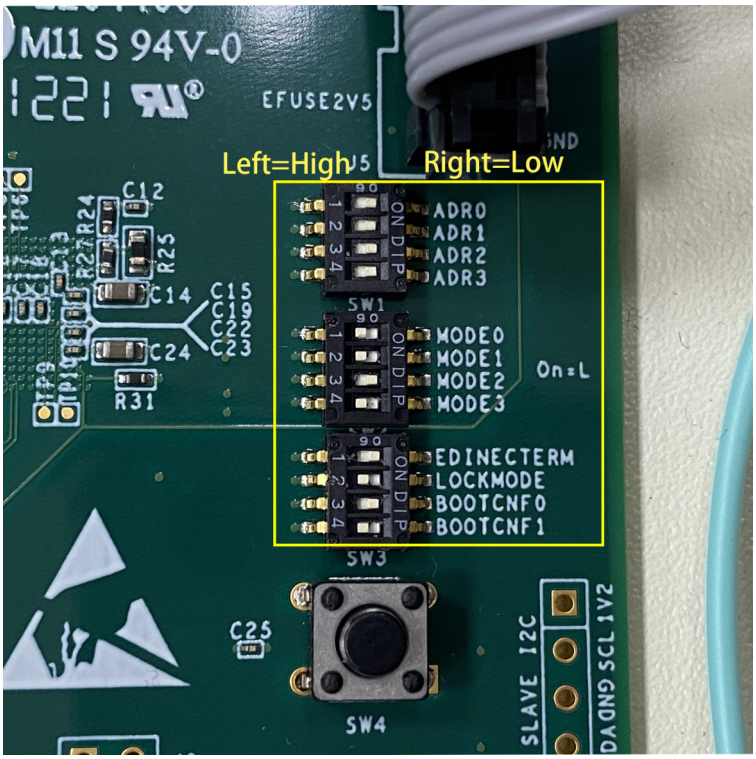
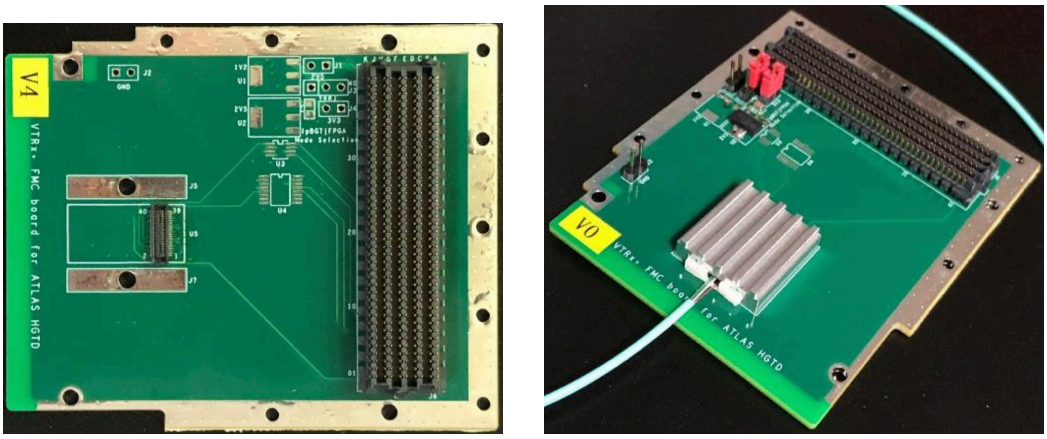


Figure 5.11: Close-up of the three dip switches.

Table 5.2: Correspondence between the mode pins settings and the lpGBT working modes.

MODE [3:0]	Tx Data Rate	Tx Encoding	lpGBT Mode
4'b0000	5.12 Gbps	FEC5	Off
4'b0001	5.12 Gbps	FEC5	Simplex Tx
4'b0010	5.12 Gbps	FEC5	Simplex Rx
4'b0011	5.12 Gbps	FEC5	Transceiver
4'b0100	5.12 Gbps	FEC12	Off
4'b0101	5.12 Gbps	FEC12	Simplex Tx
4'b0110	5.12 Gbps	FEC12	Simplex Rx
4'b0111	5.12 Gbps	FEC12	Transceiver
4'b1000	10.24 Gbps	FEC5	Off
4'b1001	10.24 Gbps	FEC5	Simplex Tx
4'b1010	10.24 Gbps	FEC5	Simplex Rx
4'b1011	10.24 Gbps	FEC5	Transceiver
4'b1100	10.24 Gbps	FEC12	Off
4'b1101	10.24 Gbps	FEC12	Simplex Tx
4'b1110	10.24 Gbps	FEC12	Simplex Rx
4'b1111	10.24 Gbps	FEC12	Transceiver

VTRx+ module. For more details about VTRx+, see Section 4.4.3.4.



(a) Without VTRx+ module and heat sink installed (b) With VTRx+ module and heat sink installed

Figure 5.12: Picture of the VTRx+ daughter board.

The connection of the VTRx+ 40 pins to the FMC connector also follows the VITA 57 standard. For example, the four high-speed uplinks and one high-speed downlink are connected to the pins specified for multi-gigabit transceiver data transmission, which allows the VTRx+ to interact with the high-speed transceivers of FPGA. The 1.2 V and 2.5 V power supply pins are connected to the pins specified for power supply, which can assure the correct voltages are delivered to the board when it's inserted into the carrier board. However, when it's inserted into a Xilinx evaluation board, the pins that should deliver 1.2 V and 2.5 V voltages may possess different voltages provided by the evaluation board. To solve the issue, two LDO are used to regulate the possibly-mismatched voltages to the target voltages (1.2 V and 2.5 V). In addition to the power pins, signal pins like VTRx_DIS, VTRx_RSTN, VTRx_SCL and VTRx_SDA are also connected with FPGA pins, since the IO voltage level on the VTRx+ side may mismatch the IO voltage level on the FPGA side, two level translation ASICs are used for respectively handling the I^2C signals and the other two signals. The RSSI pin outputs current that is proportional to the received signal strength of the VTRx+ transimpedance amplifier (TIA). Therefore the pin is pulled up by a resistor to 2.5 V and pulled down by two resistors, and the voltage between the two resistors is routed to the FMC connector and finally goes to the ADC[1] of the DAQ lpGBT. Finally, there are two thermistor sensor pins, one is pulled up by a resistor to 1.2 V and routed to the FMC connector, at last reaches the ADC[2] of the DAQ lpGBT, the other is not used and just grounded.

It's found that the VTRx+ RSSI got drifted down with time (the famous VTRx+ failure) in some LHC experiments. After years of investigation, it's concluded the main cause of this issue is the poor heat dissipation of the device. With good thermal design, that failure is expected to be avoided. For our case, we adopt the VTRx+ retaining clip design from

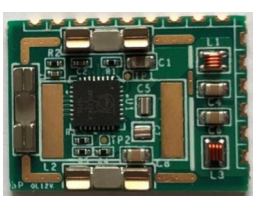
CERN and a custom-made metallic heat sink. To install those components, the VTRx+ module is firstly assembled with the clip, then it's inserted into the electrical connector, finally the heat sink covers the entire VTRx+ module from above and gets fixed to the board by M2 screws. Figure 5.12(b) is a picture of the board with fully-equipped VTRx+ mounted.

The substitute board of the VTRx+ daughter board is a SFP+ mezzanine card, which is described in Section 5.2.7.

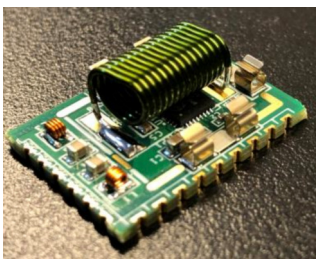
5.3.2.3 bPOL12V daughter boards

The bPOL12V daughter board for modular PEB uses the bPOL12V version 4 as the main working device, and it aims to route all the important pins of the ASIC out for easily interfacing with the carrier board. Considering the fact that this board is also targeting for the PEB, it has to be designed with as small profile as possible both in board area and thickness. The design of the bPOL12V ASIC is being actively iterated by the CERN DCDC project group, and the bPOL12V version 6 was released soon after the version 4, so the daughter board has also been updated accordingly. For more details about bPOL12V, see Section 4.4.3.3.

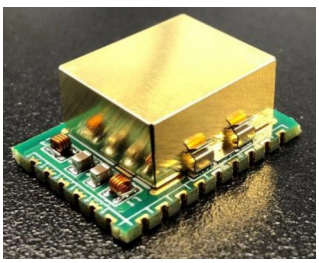
Figure 5.13(a) shows the top view of the daughter board. It is a rather small board with the size of $26\ cm \times 19\ mm$. The only ASIC is mounted on the middle area and surrounded by some auxiliary components like resistors, capacitors and inductors. Hence, its design is quite straightforward, and it is a four-layer PCB with top and bottom layers used for wire routing and middle two layers used for the grounds. To avoid thickness contribution from the height of connectors, the board adopts stamp hole design. As shown on the picture, the upper and right edges of the board are taken by those electrical semicircles (so-called stamp hole), which allow the board to have a direct physical and electrical contact with the carrier board by just soldering the semicircles on both boards together. For a convenient and efficient operation, during the test of the modular PEB system, pin header connectors are soldered to the semicircles for handy installation and disassembly to/from the carrier board.



(a) Top view



(b) With air-core inductor



(c) With shielding case

Figure 5.13: Pictures about the bPOL12V daughter board.

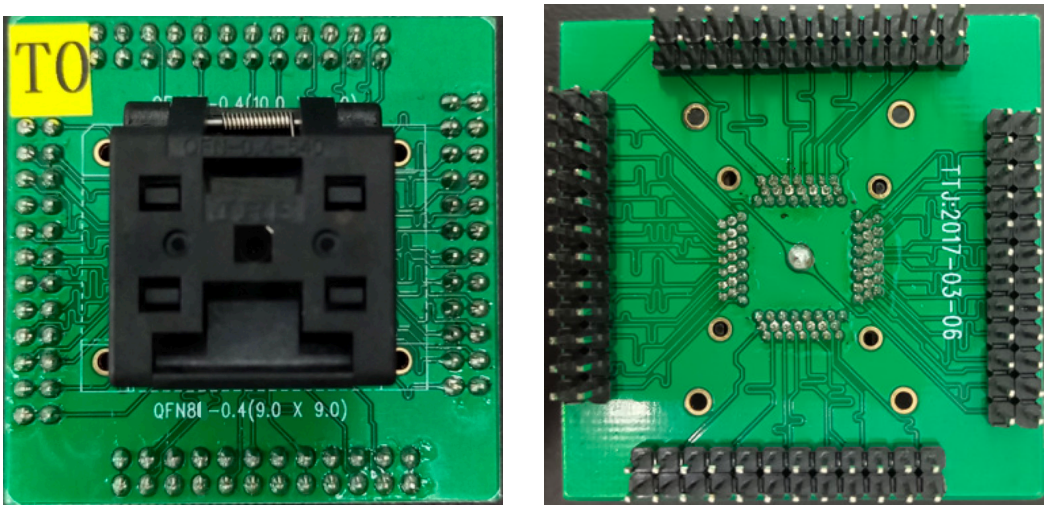
The pins are managed based on the recommendations from the ASIC data sheet and PCB design guidelines [146]. An air-core inductor with the inductance of 500 nH is used in series between the Vout and Phase pins, which allows an optimal switching frequency of about 1.5 MHz. This frequency value is set by placing a 15 $k\Omega$ resistor between the frequency selection pin Rf and the board ground. A capacitor of 200 nF is used between the bootstrap and Phase pins to provide the transient current to the high side (HS) drivers during switching, and it has to be positioned as close as possible to the ASIC footprint. The Vi pin is provided with a feedback voltage of 629 mV, which is originated from the Vout pin with a two-resistor voltage divider, and it's required that the resistor between the Vi and Vout pins should be 500 $k\Omega$, thus the value of another resistor is determined and should be 523 $k\Omega$. The pins Bit1_Vout and Bit2_Vout are used to fine-tune the output voltage based on the nominal output value by +6.67%, −6.67% and −13.3%, which may help solve the module power supply issue caused by the voltage drops on the FLEX cables with diverse lengths. All the ground pins are connected to a large power plane. The thermal pad of the ASIC is soldered to the board and also connected to ground by a large number of vias. Most importantly, the pins for input and output voltages should be connected with filter elements like air-core inductors and capacitors. Finally, all the key pins are routed to the semicircles, including the pins of input voltage, output voltage, temperature sensor, power enable, power good, and the two pin for output voltage fine-tuning.

As we know, the larger inductance the inductor has, the bigger size it will be. With such an inductance of 500 nH, the air-core inductor is expected to be very space-occupying. As can be seen in Figure 5.13(a), the two large golden pads on both sides of the ASIC are reserved for the air-core inductor, which even straddles the entire ASIC and other components. Two types of air-core conductors are successively used, one is the 2929SQ-501 from the Coilcraft company and it has the height 7.24 mm, the other one is the S15100075 from Wurth company and its height is 3.3 mm. The latter is a better choice since it can keep the board with a small thickness. In addition to these two, other types of inductors are also been investigated in parallel. Figure 5.13(b) shows a picture of the bPOL12V daughter board with the higher air-core inductor mounted. The board is sensitive to external magnetic field, plus the HGTD will be working under the magnetic field environment, therefore a shielding design has to be considered for this board. A shielding case with the profile of 18 mm × 15 mm × 8 mm is designed to cover the board, and it uses the 200 μm copper-zinc alloy for the consideration of cost at the moment, and in the future it can be changed to other materials with low material budget such as the 100 μm gold-plated aluminum adopted by the ITK EoS DCDC converter, but the ferro magnetic materials like iron, cobalt, nickel and manganese should be avoided. As can be seen from Figure 5.13(a), three SMT shield clips are mounted on the three sides of the board for holding the shielding case. This kind of design can facilitate possible repairs of the board during tests. Figure 5.13(c) shows a picture of the bPOL12V

daughter board covered by the shielding case.

5.3.2.4 MUX64 daughter board

Similarly, the MUX64 daughter board is also designed to route all the pins of MUX64 to the pin header connectors for easily interfacing with the carrier board. Figure 5.14 shows the top and bottom views of the board. As shown in the pictures, this is a pretty simple board with just one components on the top and pin header connectors on the bottom.



(a) Top view (b) Bottom view

Figure 5.14: Pictures of the MUX64 daughter board.

This component on the top is actually a test socket with the QFN88 (quad flat no-lead) package and houses the MUX64 die. The motivation of using the test socket is that, at the time of this board design, the bare dies of MUX64 were just delivered to us and a lot of tests are required to be done before they get packaged with QFN88. Inside the test socket, there is one slot dedicated for the MUX64 bare die and the slot is wire-bonded with 25 μm gold wires to its QFN88 pins. By just placing the bare die into the slot and closing the lid, each pin of the ASIC can have a good electrical contact with the socket. With the help of the socket, those bare dies can be easily mounted and unmounted to/from the daughter board without soldering and desoldering work, therefore they can be tested even with just one such kind of daughter board. Afterwards, with the progress of the MUX64 die tests, those passing the tests are qualified to be packaged. Hence, a new version of the MUX64 daughter board is designed by using the QFN88-packaged MUX64, but still adopting the same pin header connectors and pin assignment. No matter which version of the board is used on the modular PEB, it should be properly working in the same way with the other version.

The connectivity from the MUX64 to the pin header connectors is quite straightforward. The 64 signal input pins, six channel selection pins and one output pin are directly connected

to the connectors, and several capacitors with different capacitances are used in parallel with the 1.2 V voltage supply pins for the purpose of power filtering.

5.3.2.5 FH26W daughter board

The FH26W daughter board is an interface board used for re-routing the signals from the module connectors (FH26W series from Hirose company) to FMC connector. On one side, five modules are attached to the FH26W connectors of the board through FLEX cables. On the other side, the board can be easily integrated into the carrier board through the FMC connector, thus the modules are connected to the carrier board in such an indirect way. Figure 5.15 shows a top view of the board. All the components used on the board are connectors and they are all placed on the top side. Those connectors include five 71-pin FH26W connectors respectively for the signals and LV supply of five modules, five 13-pin FH26W connectors respectively for the five modules HV supply, five SMA connectors on the upper side for the five modules HV inputs, one FMC connector on the right side for interfacing with the carrier board, 15 pin header connectors on the bottom side for setting the three LSBs of each module.

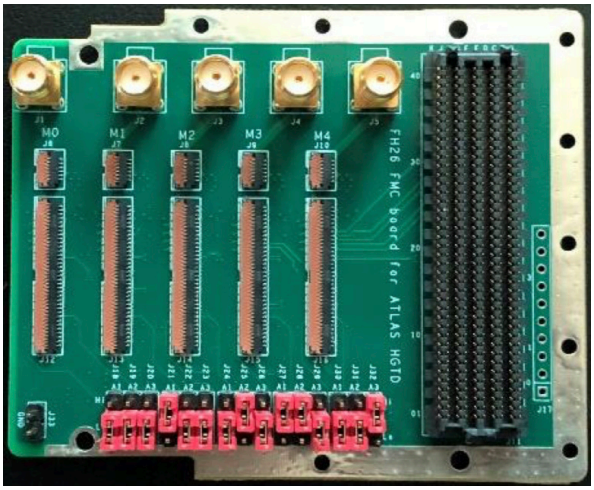


Figure 5.15: Top view of the FH26W daughter board.

The FH26W series connectors are selected as the baseline connectors for the PEB design, and their feasibility and reliability can be evaluated through this board. In addition, as mentioned in Section 5.1, the distance between two adjacent connectors can also be evaluated, a tentative value of 10 mm is used in this design. The connectors are actually high-density FPC connectors with 0.3 mm pitch and 1.0 mm height. They have two rows of pins with staggered arrangement such that one row has even number of pins while the other row has odd number of pins. The 13-pin connector has a profile of just 5.4 $mm \times 3.5$ mm while the 71-pin connector has a profile of 22.8 $mm \times 3.5$ mm. They have to be handled with care due to its thin and small design. In terms of FPC insertion, the flip lock mechanism is easier to operate and works with a light force. A clear tactile click of the actuator confirms the

completion of the mating process. In terms of FPC removal, just slightly pull it out after releasing the lock by lifting up the actuator. The HV isolation performance of two neighbouring HV connector pins are extremely crucial for the HV supply. It's already tested that the breakdown voltage (also called dielectric withstanding voltage) can reach up to $2400\ V$ with the leakage current less than $1\ \mu A$ under both the room temperature and the low temperature of $-30^{\circ}C$. The breakdown voltage exactly satisfies the requirement, which is three times the maximum working voltage ($800\ V$).

In terms of the connectivity among those connectors, the 71-pin connectors are just considered to be routed to the FMC connector while the 13-pin HV connectors are routed to their corresponding SMA connectors. The signals on each 71-pin connector include timing and luminosity data, clock, fast command, slow control, analog and digital LV supply, grounds, etc. All of them from the five connectors should be properly connected to the FMC connector, which brings complexities to the wire routing of the board. For the pin assignment of the FMC connector, like the other types of daughter board, it follows exactly the same pin assignment with that of its counterpart FMC connector on the carrier board. More specifically, it also adopts the VITA 57 standard to the largest extent to be able to compatible with the Xilinx evaluation board. For example, the analog and digital power supply of the first and second modules are connected to the pins for power supply, therefore the connector slots for them can be used for module test when the board is integrated into Xilinx evaluation board. When it comes to module emulator, all the five slots can be used since the module emulator can also be powered up through a dedicated connector on its board. Other signals are just connected to the pins for user data with quite large freedom, and those user data pins are associated with the FPGA configurable I/Os thus can satisfy the requirement of those signals.

For the pin assignment of the FH26W connectors, it should match with that of the connectors on the module side. Those pin assignments are already determined in the early days right after the FH26W series connectors are selected as the baseline connectors, since they significantly affect the designs of not only this board but also the module FLEX and FLEX cable(or tail). In addition, the center-to-center distance between the 13-pin and 71-pin connectors is chosen to be $15.6\ mm$. Figure 5.16 shows the pin assignments of the FH26W connectors on both the module and interface board sides. It's worth mentioning that the FLEX cable should be inserted to the connectors with its contacts facing downwards.

5.3.3 Auxiliary boards

5.3.3.1 Module emulator board

The module emulator board is designed to mimic the real HGTD module both in dimension and functionality since the module is not expected to be ready at the stage of the modular

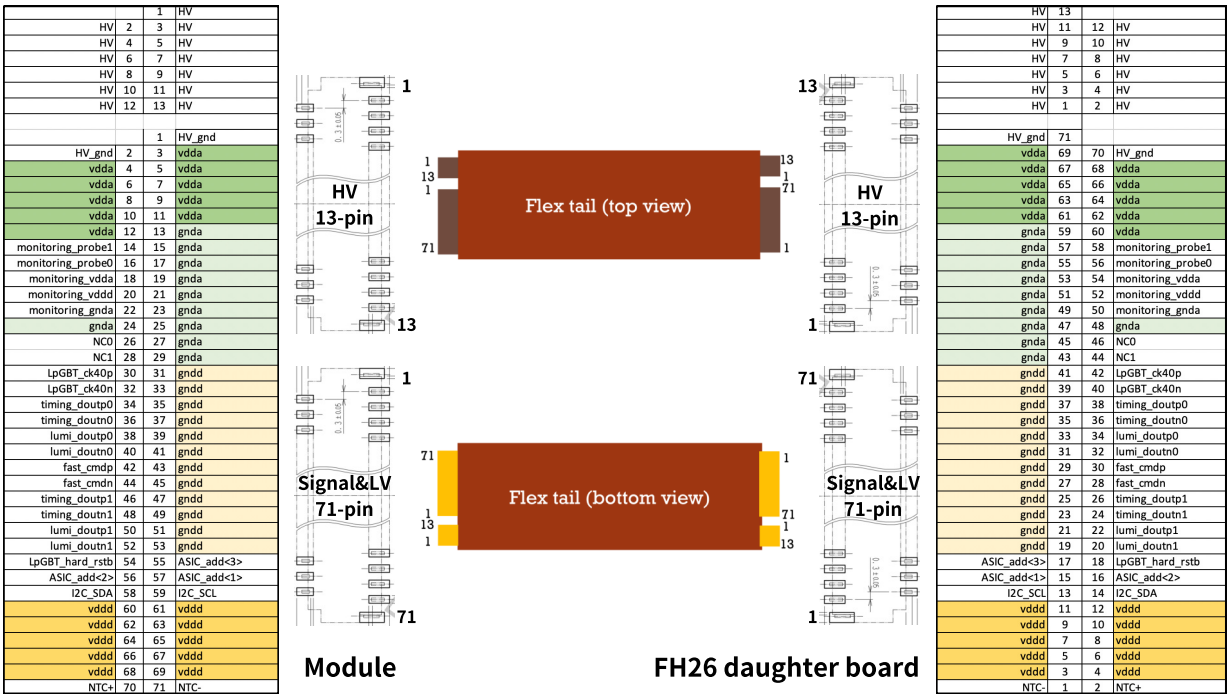
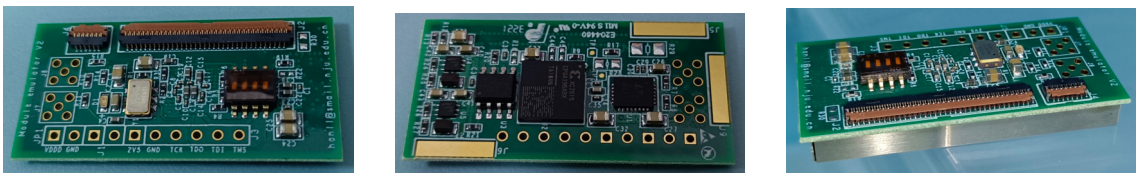


Figure 5.16: Pin assignments of the FH26W connectors on both the module and FH26 daughter board sides.

PEB.

5.3.3.1.1 Hardware structure The idea is that the board uses FPGA to realize the ALTIROC logic and interfaces with the FH26 daughter board by using the same FH26W connectors with the real module. This board goes through two versions. Compared to the first version, the second version is designed with the latest mechanical parameters (including board size, center-to-center distance between FH26W connectors, connector-to-edge distance), the new power supply of the FPGA bank for signal, the updated pin assignment and shielding case mounting method. Figure 5.17 shows some pictures of the module emulator board. Like the module, the board has a small profile of 40 mm × 19.8 mm, where the center-to-center distance between FH26W connectors is 15.6 mm while the distance between the connector center and the board edge is 3 mm.



(a) Top view (b) Bottom view (c) With shielding case

Figure 5.17: Pictures of the module emulator board.

As can be seen from the pictures, both the top and bottom are almost occupied by components. To mimic the glue dots area of the module, 4 mm width on left and right sides

of the top should not be placed with any components, which makes the board surface even more crowded. Therefore, the wire-routing of the board is one of the most challenging parts of the design. In order to have a proper wire-routing, the board adopts the eight-layer stack-ups, among which up to four layers are used for wire routing, two layers for power planes and two layers for ground planes. Another way of mitigating the congested situation is to select the components with smaller footprints, for example, instead of using the bigger 0603 package ($1.55\ mm \times 0.85\ mm \times 0.45\ mm$) that is beneficial to maintenance and replacement, the smaller 0402 package ($1.0\ mm \times 0.5\ mm \times 0.35\ mm$) is used as much as possible for resistors and capacitors to save space.

On the top side (Figure 5.17(a)), there are just those auxiliary components, such as the two FH26W connectors on the upper side, the two micro-miniature coaxial (MMCX) connectors on the left side for the time domain reflectometer (TDR) measurement of the FLEX cable transmission lines, the pin header connectors on the lower side for FPGA programming, power supply and testing purposes, the dip switch on the right side for reset and other external settings, the shiny crystal oscillator in the middle area for providing external clock of the FPGA, and some other resistors and capacitors. On the bottom side (fig:modularpeb:hwdesign:hwmev2:bottom), there are all the main ASICs, including the biggest ASIC (FPGA) in the middle, the FLASH memory ASIC on the right of the FPGA for storing the bitstream file of the FPGA, the LDO regulator and two DCDC switching regulators on the left of the FPGA for the voltage conversions of the module emulator, also some resistors, capacitors and inductors distributed around those ASICs. Taking into account the test of the module emulator, a shielding case is also designed to cover the entire bottom side of the board, aiming to serve as support structure of the board and provide a flat and stable pedestal and proper heat dissipation (with the help of thermal conductive silicone used in the gap between the shielding case and the FPGA) when the board gets tested individually or in a stack-up way. The shielding case, made of $0.2\ mm$ copper-nickel alloy, has a slightly smaller size of $38.3\ mm \times 17.5\ mm$ compare with the board size and a height of $4\ mm$. The mounting strategy is exactly the same with that of bPOL12V daughter board described in Section 5.3.2.3. As can be seen on the bottom side, there are three golden pads on the edges, which are used for soldering the shield clips that hold the shielding case. Figure 5.17(c) shows a picture of the module emulator with the shielding case mounted on the bottom side.

5.3.3.1.2 FPGA selection The FPGA for the module emulator is selected among all the Xilinx 7-series FPGAs, including Spartan-7, Artix-7, Kintex-7 and Virtex-7. There are three main points considered during the selection. Firstly, the size of the FPGA package should be as small as possible given the very limited space on the board. Secondly, the FPGA resources, such as the number of logic cells, I/O pins, RAM (random access memory) blocks

and DSP (digital signal processor) slices, should be adequate for the implementation of the ALTIROC2 digital functions. Thirdly, the FPGA should have a low cost since there will be a large amount of module emulators needed by several modular PEB setups distributed on different sites. After careful consideration, the Spartan-7 FPGA XC7S15-2CPGA196C is finally chosen. It uses BGA package with the size of $8\ mm \times 8\ mm$, and it has 14×14 pins with the pitch of $0.5\ mm$ and up to 100 user I/O pins.

The I/Os in 7 series FPGAs are classified into high range (HR) and high performance (HP) banks [154], where the HR I/Os offer the widest range of voltage support, from $1.2\ V$ to $3.3\ V$, while the HP I/Os are optimized for highest performance operation, from $1.2\ V$ to $1.8\ V$. The selected FPGA has two HR banks [155], bank 34 and bank 14, which can be supplied with two different voltages for supporting different I/O standards of the signals connected to the banks. Each bank has up to 50 user I/O pins that can be used in either single-ended or differential way, and also some dedicated pins for the power supply of the I/O pins. The separate power supply for I/O pins allows multiple I/O standards used in the same FPGA and flexible I/O standard selection for a single bank. There are several tens of I/O standards supported by the FPGA [156], such as the LVDS (low voltage differential signaling), LVCMOS (low voltage CMOS), HSTL (high-speed transceiver logic), SSTL (stub-series terminated logic), etc. In addition to the user I/O banks, there is also a bank 0 where dedicated configuration pins are located. Furthermore, the I/O pins of the bank 14 can also be working as configuration pins [157]. Therefore, the FLASH memory, programming interface and configuration mode selection can be managed within the bank 0 and bank 14 of the FPGA. Also other types of power supply pins, such as VCCAUX pins for auxiliary circuits ($1.8\ V$), VCCAUX_IO pins for auxiliary I/O circuits ($1.8\ V$ or $2.0\ V$), VCCINT pins for the internal core logic ($0.9\ V$ or $1.0\ V$), VCCBRAM pins for the FPGA logic block RAM ($1.0\ V$), must be properly handled.

5.3.3.1.3 I/O standard selection The I/O standard match between the module emulator side and the modular PEB side is crucial for a robust and reliable data transmission. On the modular PEB side, if FPGA would be used as the substitute of the lpGBT, the match can be easily achievable by configuring the FPGA pins with proper I/O standard on both sides; If the lpGBT would be used, an appropriate I/O standard has to be chosen for the FPGA pins of the module emulator to be able to match the specific I/O standard called CERN low power signaling (CLPS) adopted by the lpGBT. Therefore, the idea is to find an optimal one among all the I/O standards supported by the FPGA.

Figure 5.18 shows a general overview of the CLPS standard. As can be seen, the digital signal is transmitted in the form of square wave, and its voltage level swings around the common mode voltage. When the signal is fed into the lpGBT as input, it's required that, the high level voltage is larger than $0.9\ V$ and the low level voltage is less than $0.3\ V$

in case of single-ended transmission, the common mode voltage should be within 0.07 – 1.13 V and the differential voltage can vary from 0.14 V to 0.45 V in case of differential transmission. When the signal is driven out of the lpGBT, the high level voltage is 1.07 V or 1.18 V and the low level voltage is 0.13 V or 0.02 V in case of single-ended transmission, the common mode voltage is 0.6 V and the differential voltage can vary from 0.1 V to 0.4 V (under the condition of a 100 Ohm differential termination impedance) in case of differential transmission. Therefore, the FPGA pins of the module emulator that receive the signals from the lpGBT should be set with proper I/O standard such that they can match the electrical characteristics of the CLPS output signal, similarly, the FPGA pins of the module emulator that send signals to the lpGBT should also be set with proper I/O standard such that they can generate suitable signals that satisfy the requirements demanded by the CLPS input.

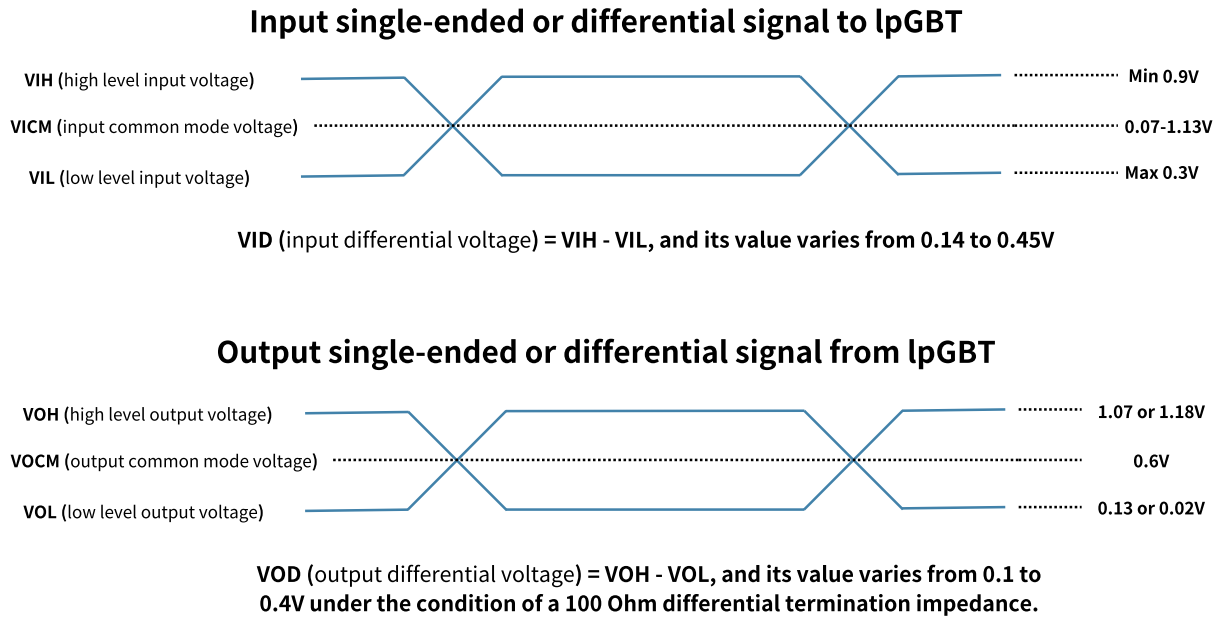


Figure 5.18: CLPS standard introduction in the cases of single-ended and differential signals as input and output.

In terms of the I/O standard selection for single-ended signal, it is quite straightforward. The idea is to find the suitable I/O standards from the user guide [158] that describes the DC input and output levels of the FPGA standard with the single-ended CLPS signal features in mind. Especially, those parameters (VIL, VIH, VOL and VOH) are the key indicators for selecting the satisfactory standards. The good news is that there are several standards that can fulfill the requirements, such as the LVCMOS12 and HSUL_12.

In terms of the the I/O standard selection for differential signal, there are more aspects that need to be considered. In addition to the electrical parameters of VICM, VID, VOCM and VOD, the transmission performance based on the certain I/O standard should be also

taken into account. As we know, the most commonly-used I/O standard in the FPGA is the LVDS, also called LVDS_25. The pins with this standard can receive CLPS output signals, but the VOCM of this standard ($1.25\ V$) is beyond the common mode voltage range of the CLPS input ($0.07 - 1.13\ V$). Actually, there is a workaround for the mismatch between the LVDS_25 output and the CLPS input. It is called AC-coupled and DC-biased method, which requires two capacitors in series with the differential pins on the module emulator and two-resistor voltage divider for providing proper bias voltages that CLPS input can accept to the differential pins of the lpGBT. However, with this method, there must be extra capacitors and resistors used on both module emulator and modular PEB, which makes the crowded layout and challenging wire routing issues even worse. Therefore, it's decided that we drop this method and pick up new candidates from other I/O standards.

There are about fifteen differential I/O standards supported by the HR bank, such as the DIFF_HSTL_I, DIFF_HSUL_12, DIFF_MOBILE_DDR, DIFF_SSTL15, DIFF_SSTL18_I, etc. As input, all of them can support the CLPS output signal based on the VICM and VID provided by the user guide, but as output, VOCM and VOD of most standards are not specified clearly in the document. An experiment was setup to measure the VOCM and VOD of those standards. The idea is that, two FPGA differential pins are firstly set with a specific I/O standard, then the output of one of the two pins is observed with oscilloscope, finally the VOCM and VOD can be determined from the waveform. It's worth mentioning that the FPGA used in the test board is also a 7-series Xilinx FPGA XC7K325T and the two differential pins of the FPGA are terminated with a $100\ \Omega$ resistor for the purpose of impedance match. Figure 5.19 shows an example of the waveform measured from the differential pins with the LVDS_25 standard, where the horizontal arrow indicates its VOCM ($1.27\ V$) while the vertical arrow indicates its VOD ($0.43\ V$). Clearly, the VOCM is beyond the range of the CLPS VICM, so LVDS_25 is not chosen as the candidates. Other I/O standards are also measured using the same test strategy. Table 5.3 summarizes the measurement results of those I/O standards. By looking at the values of VOCM, most of them are within the range of the CLPS VICM. Further considering the swings (i.e. VOD) around the VOCM, those VOCM values that are close to the range edge may have a worse performance than those in the middle of the range. Therefore, more I/O standards get ruled out based on the consideration, at the end, there are just a few standards left, including PPDS_25, DIFF_SSTL135, DIFF_SSTL135_R and DIFF_HSUL_12.

Among the few candidates, the one with the best transmission performance should be further picked out. To do that, another experiment is set up to measure the bit error rate (BER) of the data transmission under the use of a certain I/O standard. Figure 5.20 shows the block diagram of the test setup. Starting from the pseudo random bit sequence (PRBS) generator, the random data is firstly generated and fed into the serializer, which serializes the data to the speed of $1.28\ Gbit/s$. Through the differential pins set with a specific I/O

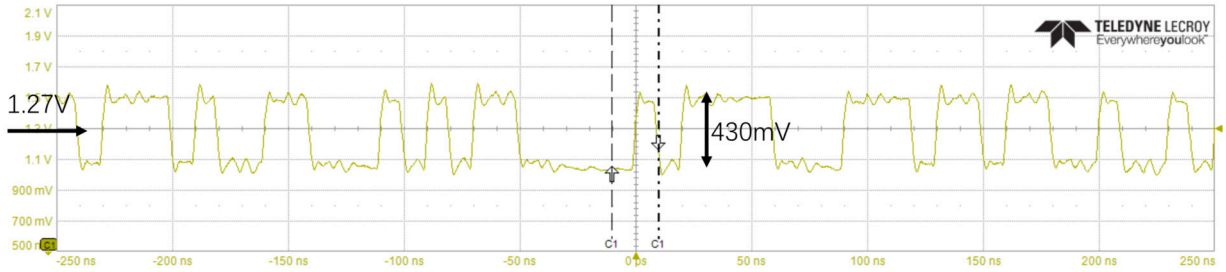


Figure 5.19: The waveform measured from the differential pins with the LVDS_25 standard.

Table 5.3: The measured VOCM and VOD of different I/O standards.

I/O standards	VOCM	VOD
LVDS_25	1.27 V	0.43 V
PPDS_25	0.66 V	0.21 V
DIFF_HSTL_I_18	0.93 V	1.3 V
DIFF_HSTL_II_18	0.90 V	1.4 V
DIFF_SSTL_18_I	0.93 V	1.2 V
DIFF_SSTL_18_II	0.96 V	1.4 V
DIFF_MOBILE_DDR	0.85 V	1.2 V
DIFF_SSTL15	0.76 V	1.1 V
DIFF_SSTL15_R	0.72 V	1.0 V
DIFF_HSTL_I	0.70 V	1.0 V
DIFF_HSTL_II	0.73 V	1.1 V
DIFF_SSTL135	0.65 V	1.0 V
DIFF_SSTL135_R	0.64 V	0.94 V
DIFF_HSUL_12	0.59 V	0.87 V

standard such as DIFF_HSUL_12, the data stream is sent out from the transmitter FPGA to the receiver FPGA where a deserializer based on the Xilinx IP core XAPP1017 [159] is used to decode the data stream into original PRBS values. Following the deserializer, the error checker and error counter are used to check the correctness of the recovered PRBS values and count the number of wrong bits that will be displayed to the user with the help of the FPGA integrated logic analyzer (ILA). Finally, the BER can be calculated as the ratio of the number of wrong bits to the total number of bits received by the FPGA. It's found that the I/O standard DIFF_HSUL_12 gives the lowest BER among all the candidates. Therefore, the DIFF_HSUL_12 should be the first choice in terms of I/O standard selection for the signals on the module emulator side.

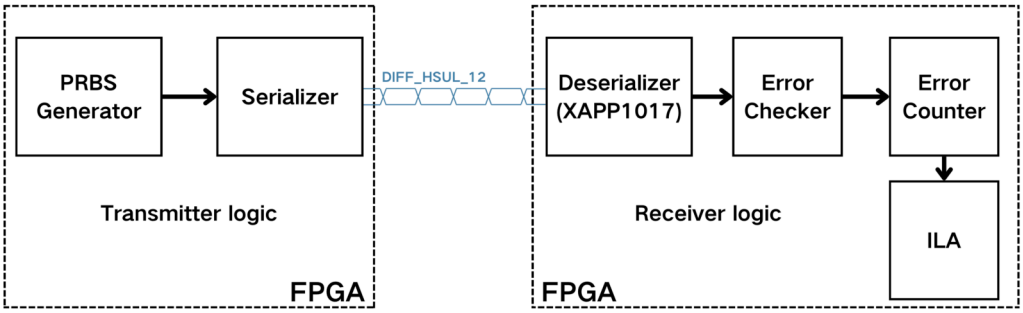


Figure 5.20: Block diagram of the setup for the BER measurement of different I/O standard candidates.

5.3.3.1.4 Clock The module emulator has two clock sources, which are the on-board oscillator clock and the lpGBT clock. Both of them will go to the dedicated MRCC (multi-region clock capable) pins of the FPGA so that the clocks can be routed to multiple clock regions. Once the clocks are delivered to the FPGA, they will be processed with the clock IP core MMCM (mixed-mode clock manager) [160], and many different types of customized clocks can be simultaneously regulated according to the clock parameters set by users thus support all the other logics implemented on the FPGA. For the on-board oscillator clock, it has a differential output with the frequency of 200 MHz and the differential pins adopt the LVDS_25 standard, then they are connected to the MRCC pins of the bank 14 whose power supply should be 2.5 V so that the MRCC pins can also be configured with the same I/O standard with the input clock. For the lpGBT clock, it is also a differential clock but from the lpGBT E-ports, and it adopts the CLPS standard. Hence, the lpGBT clock is connected to the MRCC pins of another bank (bank 34) supplied with the voltage of 1.2 V, and the MRCC pins are configured with the I/O standard DIFF_HSUL_12. In addition, the lpGBT clock is highly-configurable and a suitable frequency (40 MHz or 320 MHz) can be set based on the requirements of the future test. For example, the ALTIROC version 2 requires a 40 MHz clock from lpGBT while the ALTIROC version 3 requires a 320 MHz

clock from lpGBT. Depending on which ALTIROC version will be emulated, the frequency of the clock should be set accordingly.

Figure 5.21 shows a the schematic of the clock design on the module emulator. As can be seen, the MMCM is the crucial part for the clock generation and distribution inside the FPGA. It can be instantiated through the dedicated graphical user interface (GUI) called clocking wizard from the host computer. There are many clock parameters that can be set through the GUI in a customized way, which makes the clock generation more adjustable. For example, the way the clock is regulated can be chosen from the MMCM and PLL, and both of them serve as frequency synthesizers for a wide range of frequencies, jitter filters for either external or internal clocks and the resources for deskewing clocks, but the PLL is just a subset of the MMCM functionality. Also the number of output clocks and the parameters for each of them can be easily set, and the configurable parameters include the requested output frequency, phase and duty cycle. It's worth mentioning that several MMCMs can be instantiated in one FPGA, and the number of MMCMs that can be instantiated depends on the available MMCM resources.

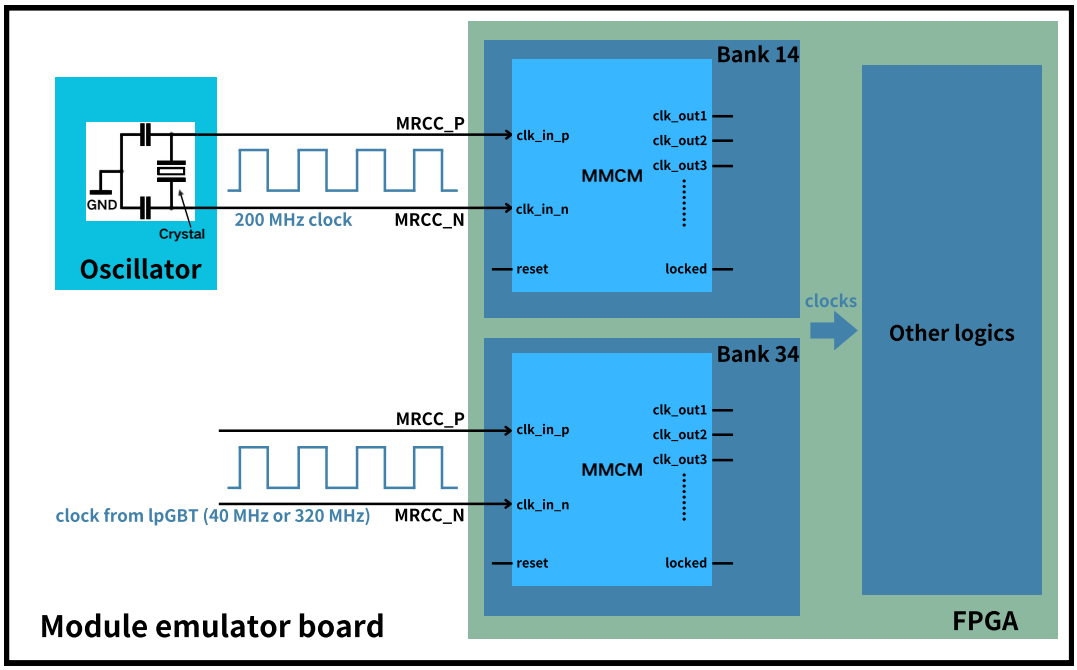


Figure 5.21: Schematic of the clock design on the module emulator.

5.3.3.1.5 Power supply The power supply of the module emulator is mainly about LV power supply since the HV power supply is just delivered to the board but not applied on any components. The HV is supplied through the 13-pin FH26W connector and its return ground (HV_gnd) is connected to the dedicated pins of the 71-pin FH26W connector. Between the HV and HV_gnd, a large resistor with the resistance of 1 $M\Omega$ and the maximum working voltage of 800 V is used in series. The LV is used for supplying all the components.

Figure 5.22 shows the block diagram of the LV power supply design.

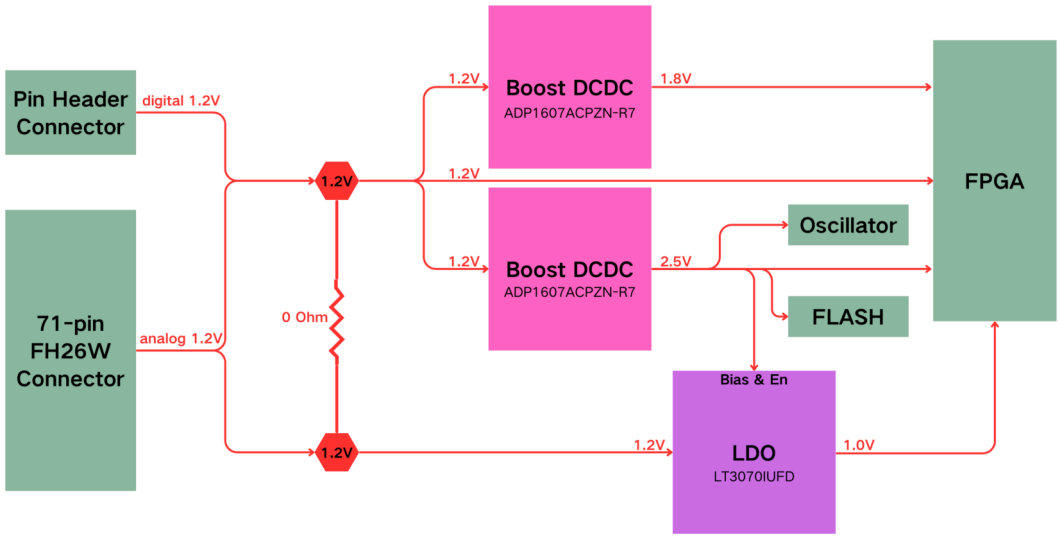


Figure 5.22: Block diagram of the LV power supply design on the module emulator.

On the module emulator board, the digital and analog power supply channels are connected together with one 0 Ω resistor since there is no need to supply the FPGA with separate digital and analog voltages like the ALTIROC does. Correspondingly, the HV_gnd, analog and digital grounds are also connected together with two 0 Ω resistors. They can be easily separated by just desoldering the 0 Ω resistors when there is new requirement of dealing them independently. The LV has two sources, either from the on-board pin header connector or from the 71-pin FH26W connector. Both of the sources should provide the voltage of 1.2 V, for the former that is used for the debugging in the early stage, an external voltage is just used to supply the digital channel, while for the latter, both the digital and analog channels are delivered from the modular PEB side via the FLEX cables. No matter which source the LV originally comes from, it will be handled in the same way with the same sets of ASICs. The idea is that, by using two boost DCDC switching regulators, the 1.2 V voltage is firstly boosted to 1.8 V and 2.5 V voltages, then it is bucked down to 1.0 V by a LDO regulator that uses the already-regulated 2.5 V as its bias and enable voltage. Hence, all the other components on board can be supplied properly by those regulated voltages.

It's worth mentioning that, during the debugging of the module emulator, the pin header connector will be temporarily used. For the sake of convenience, instead of soldering pin header connector on the board, a customized clip will be attached to the dedicated slots for providing external LV supply, and the clip can be easily removed after the debugging. The interface power clip can be seen from the Figure 5.25(b), where it is used on the right side together with the interface JTAG clip for the configuration of the FPGA on the module emulator.

5.3.3.1.6 Configuration The configuration refers to two aspects, one is how the configuration pins in the bank 0 and bank 14 are handled, the other one is the programming of the FPGA thus the implementation of specific functionality. For the former, the most important point is the setting of the three mode pins M[2:0]. In this design, they are set to be 001, which configures the FPGA to work as a SPI (serial peripheral interface) master. For the latter, a 8 *Mbit* FLASH memory ASIC (part number MX25V8035FM1I) with the interface of SPI is connected to the FPGA for the storage of its configuration file. Figure 5.23 shows the flow chart of the FPGA configuration. Every time the module emulator is powered up, in the case of the JTAG (joint test action group) interface not being occupied, the FPGA will try to load the configuration file from the memory to program the logic cells. If there is no information that was burned before on the memory or there is an update on the FPGA logic implementation, the FPGA will wait for further configuration from the JTAG interface. Then the data stream from the JTAG goes to the FPGA, and it can be configured that either the bitstream file is directly used for programming the FPGA logic cells or the configuration file is sent to the FLASH and get burned over there. The JTAG operation has higher priority over the FLASH memory loading, which means when the JTAG interface is occupied during powering up, the FPGA will not load the information from the memory.

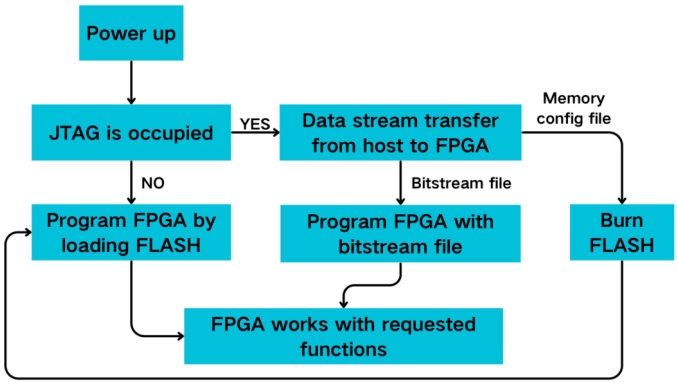


Figure 5.23: Configuration flow of the module emulator.

Figure 5.24 shows the connectivity about the FPGA configuration in the hardware aspect. As can be seen, the FLASH memory is associated with the FPGA through the SPI interface, which can support three types of communication modes (standard, dual and quad SPI modes). The SPI connection on the figure uses the quad SPI mode, where the four serial data lines can be used for data transmission. Of course, other two SPI modes can also be used during the FLASH memory burning. Apart from the four serial lines, the SPI interface also has the clock line through which the SPI master provide clock to the FLASH memory, and the chip selection line that sends the enable signal to the FLASH memory. For the JTAG interface, there are four lines dedicated for communication, including the clock line through which the JTAG programmer provide clock to the FPGA, the TMS (test mode selection) line for setting a specific test mode of JTAG, the TDI (test data input) and

TDO (test data output) lines respectively for data input and output. There is also a LED connected to the FPGA pin DONE, indicating the successful completion of configuration.

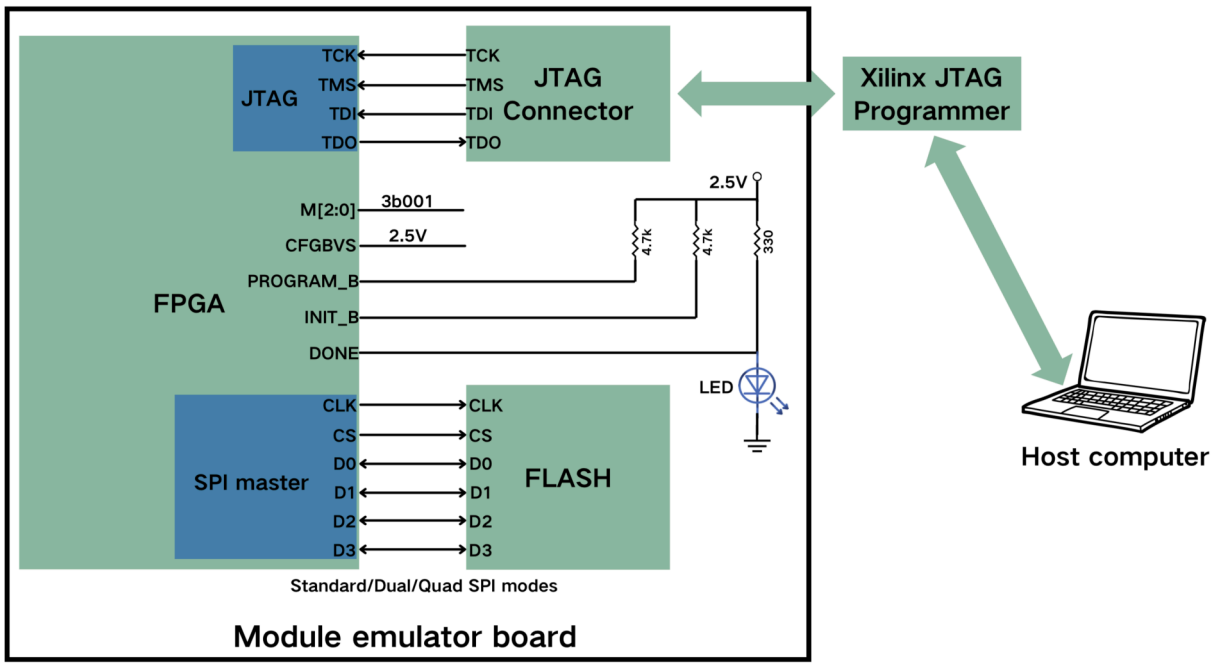


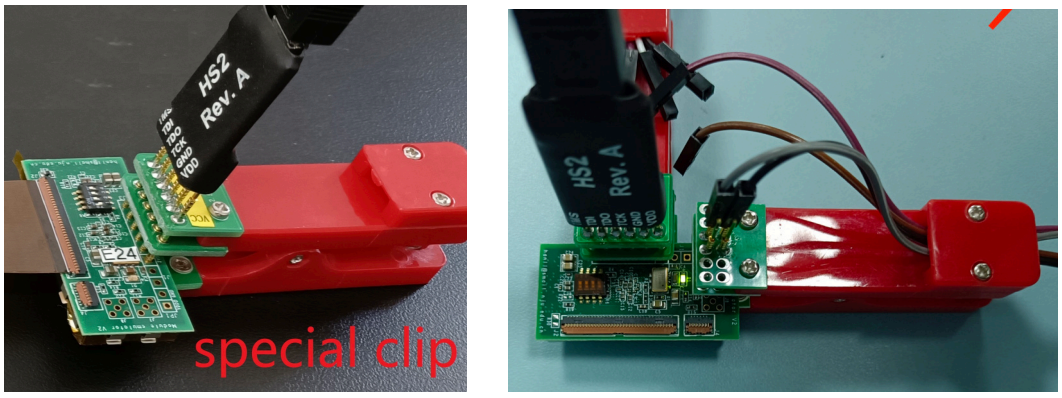
Figure 5.24: Connectivity about the FPGA configuration in the hardware aspect.

As described in 5.3.3.1.1, on the top side of the module emulator, there are slots of pin header connectors used as the JTAG interface. Due to the constraint on the emulator height, it's decided that a special interface clip is used for bridging the FPGA and the JTAG programmer instead of soldering pin header connectors to the slots. The clip can be easily removed right after the FPGA configuration. Figure 5.25(a) shows a picture of the FPGA configuration with the help of the custom-made clip. On one side, the probes of the clip should be firmly pressed down to the JTAG slots/pads on the board to make sure ohmic contact between them. On the other side, the Xilinx JTAG programmer are attached to the top of the clip.

5.3.3.2 lpGBT configuration board

In order to configure the lpGBT, a dedicated isolated USB programmer board for lpGBT (UPL) [161] was developed based on an USB-protocol converter. Compared with the already existed lpGBT programmer piGBT [150], the UPL has three main advantages, including the compact profile, cross-platform compatibility and electrical isolation.

5.3.3.2.1 Overview There will be a lot of tasks related to HGTD peripheral electronics that require lpGBT configuration all the way from the research and development stage to the installation of the PEB. Therefore, the UPL is designed to be reliable, efficient and handy so that a robust and convenient lpGBT configuration can always be achieved. In contrast,



(a) With the help of special interface JTAG clip

(b) With the help of interface clips for JTAG and power supply

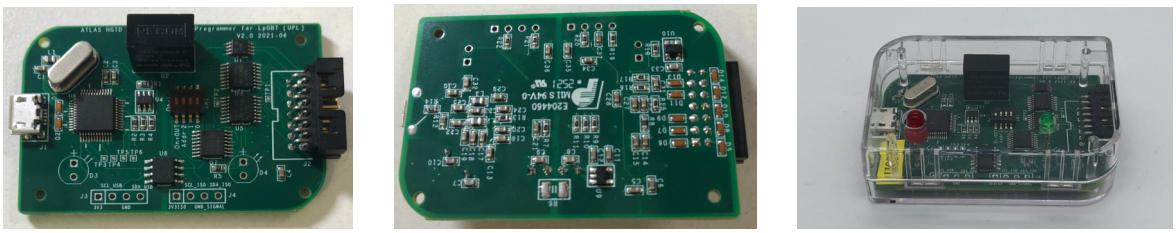
Figure 5.25: Pictures of the FPGA configuration.

the piGBT toolkit is consisted of two parts, a customized Raspberry Pi 4 for running the specific software, and a translator board for translating the 3.3 V I/O voltage to 1.2 V I/O voltage to match the lpGBT I/O standard, so it has a rather hefty size of 9 cm × 6 cm × 4 cm. It doesn't provide the electrical isolation during lpGBT configuration, which may possibly damage the lpGBT due to the surging current from the host computer.

The UPL uses a 14-pin header connector as the interface to lpGBT. Among the 14 pins, the I2C port and GPIO pins are available for the lpGBT configuration. The former is used to communicate with the lpGBT I^2C logic for updating its registers while the latter can be used to handle other lpGBT features like setting the working modes, initiating reset signal and monitoring the ready signal of lpGBT. Since the USB-protocol converter adopts the USB protocol to interface with the host side, it can be easily accessed from any computer with a specific USB driver installed, which greatly expands the platform compatibility of the UPL since basically most of the operating systems can support the USB driver, such as Windows, Linux and Mac OS. In particular, its isolation performance is completely determined by its I2C isolator and power isolator, both of which have the high isolation voltages of more than 1000 V AC RMS and 2000 V DC thus guarantees a thorough electrical isolation between the host side and lpGBT side in the aspects of power supply and data transmission. In addition, a dedicated GUI is also developed to facilitate the operation of the lpGBT configuration.

5.3.3.2.2 Hardware The UPL board is housed inside a custom-made and transparent plastic box with the dimension of 5.6 cm × 3.6 cm × 2 cm, which is way more compact than the currently-used piGBT toolkit. Figure 5.26 shows three pictures of the UPL, respectively presenting the top view, bottom view and the final product.

On the top side (Figure 5.26(a)), it can be clearly seen that there is a boundary in the middle of the board, and the boundary can also be viewed on the bottom side. This boundary is actually indicating the electrical isolation of the board. As shown from the



(a) Top view (b) Bottom view (c) Final product

Figure 5.26: Pictures of the UPL.

top, on the left side of the boundary, there are the USB-protocol converter, oscillator and EEPROM connected to the host side via a micro B USB (universal serial bus) connector. On the right side of the boundary, there are also several ASICs like GPIO expander, voltage regulator and voltage level shifters, which are associated to the lpGBT side through the 14-pin header connector. Right above the boundary, the two isolation ASICs, I^2C isolator and power isolator, are the components that make the board achieve excellent isolation performance. The power isolator placed on the upper side has the highest height among all the components, thus determining the thickness of the housing box. On the lower side, there are slots for two pin header connectors located on two sides of the boundary, one is used for debugging the I^2C signal before the power isolation while the other is used for debugging the I^2C signal after the power isolation.

On the bottom side (Figure 5.26(b)), there are just those smaller components, like filtering capacitors, pull-up resistors, ESD (electro-static discharge) suppressors and LDO regulators. From the picture of the final product (Figure 5.26(c)), there are two LEDs sticking out via the specially-reserved holes for a better view of their lights. The red one indicates ongoing communication in the USB-protocol converter while the green one indicates the readiness of the lpGBT under configuration.

Figure 5.27 shows the block diagram of the UPL in hardware aspect, using two colors to display the power supply networks before and after electrical isolation. For the power supply of the board, several LDO regulators are successively used to convert higher voltage to lower voltage. The power source comes from the host computer and gets delivered to the board through the power line of the micro B USB, and it can provide a $5\ V$ voltage with the drive current of from $100\ mA$ to $500\ mA$, which means a power of from $0.5\ W$ to $2.5\ W$ can be guaranteed for the UPL board. Then it branches out into three rails, which respectively go to the converter ASIC FT232H for its power supply, the LDO with its output voltage of $3.3\ V$ to supply the I^2C isolator, and the power isolator as its primary voltage. The isolated $5\ V$ voltage from the power isolator directly goes to the LDO, whose $3.3\ V$ output voltage is used not only to supply the GPIO expander and the high voltage side of the three level shifters, but also to support another LDO with an output voltage of $2.5\ V$ to supply the lpGBT E-fusing operation. In addition, voltage side of the three level shifters is supplied by

the 1.2 V from the lpGBT side.

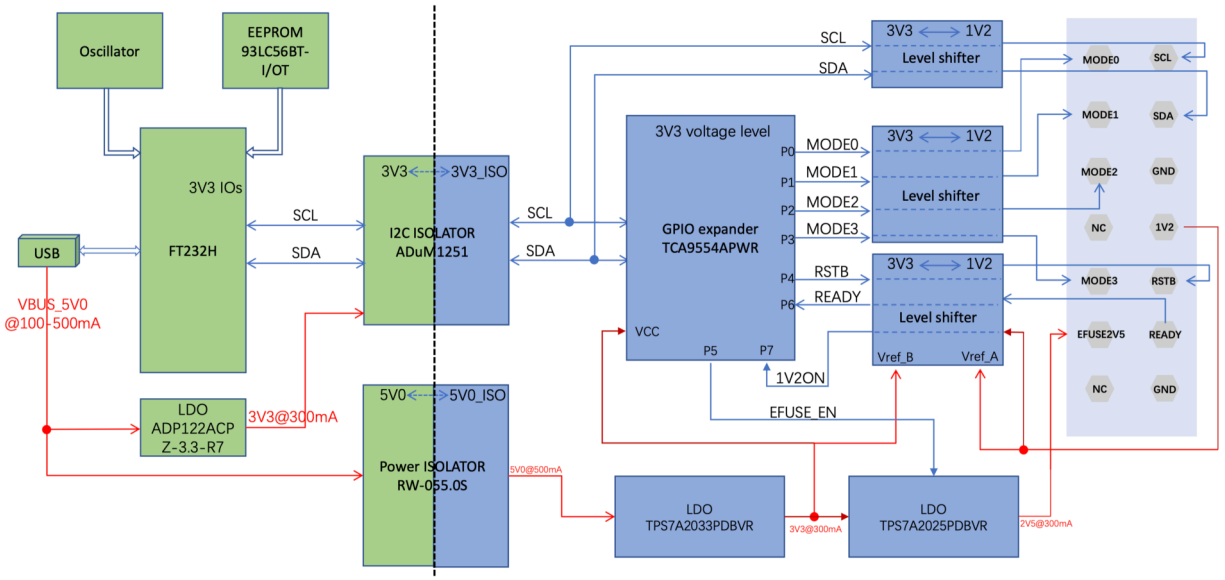


Figure 5.27: Block diagram of the UPL in the hardware aspect.

For the data transmission and communication, the host computer initiates configuration information, which is transmitted to the UPL through the micro B USB connector in the form of USB protocol. Then the FT232H converts the information of USB protocol to the data stream of I^2C protocol, which goes to the I^2C isolator and gets isolated. After the isolation, on one hand, the isolated I^2C bus, including SCL and SDA lines, goes to the level shifter to be translated to the 1.2 V voltage that lpGBT can accept; On the other hand, it is used to control the GPIO expander that has eight GPIOs, among which, seven of them are connected to another two level shifters and finally routed to the 14-pin header connector for interfacing with lpGBT while one of them is used to provide enable signal to the LDO that supplies the lpGBT E-fusing operation. In addition, an oscillator and one EEPROM are associated with the FT232H. The former provides clock to the ASIC while the latter is used to store the configuration data of it.

For the interface of UPL, there are two connectors, the micro B USB and 14-pin header connectors respectively connecting to the host side and the lpGBT side. Figure 5.28 shows a typical cable connections to this board. On one side, the custom-made flat cable is produced with the corresponding male 14-pin header connectors on both ends. On the other side, it uses the USB cable that has the micro B USB connector on one end and type A USB connector on the other end.

The pin assignment of the 14-pin header connector is shown in Table 5.4. As can be seen, it is fully dedicated to lpGBT configuration. Apart from the two I^2C pins, there are another eight pins, two of them are power pins respectively for lpGBT E-fusing power supply and lpGBT power monitoring while another six of them are GPIO pins for lpGBT I^2C control, mode setting, external reset and ready signal monitoring. Even though it's designed for

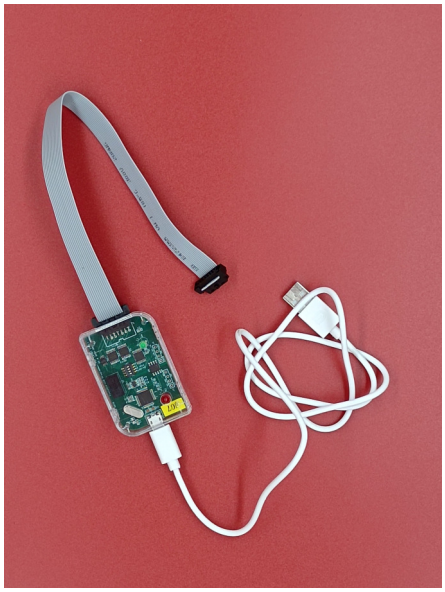


Figure 5.28: Picture of the UPL typical cable connections.

Table 5.4: Pin assignment of the 14-pin header connector of UPL.

Pin number	Name	Description
2, 4	SCL, SDA	I2C port to I2C slave of lpGBT
1, 3, 5, 9	MODE0, MODE1, MODE2, MODE3	lpGBT mode setting
8	1V2MON	lpGBT power monitoring
10	RSTB	Reset lpGBT, active low
11	EFUSE	E-fuse lpGBT
12	READY	lpGBT ready signal monitoring
6, 14	GND	Ground
7, 13	NC	Not connected

lpGBT configuration, it can also be used in other scenarios, such as configuring other I^2C slave devices with its I^2C bus and controlling other ASICs with its six GPIOs.

5.3.3.2.3 Software With about 300 configurable registers that determine the properties of its different parts, the configuration of lpGBT is quite complicated. After powering on, the lpGBT keeps looping over the states of its built-in power-up Finite State Machine (FSM). Only when those crucial registers are well configured, the FSM will jump to the last state indicating the completion of lpGBT initialization and configuration, thus the lpGBT is ready and operational. Therefore, it is of great importance to set those configuration registers properly based on the experiment requirement. In order to facilitate the lpGBT configuration, a dedicated GUI is developed based on the specific USB driver provided by the FTDI company. Figure 5.29 shows the architecture of the software design. It is consisted of two parts, the underlying blocks for directly interacting with the hardware layer and the high-level blocks for interfacing with the user side. More specifically, the high-level blocks divided into two blocks, the lpGBT configuration GUI block aims to realize

the graphic interface while the I^2C communication function block is used to implement the communication protocol. The underlying blocks are provided by the supplier of the USB-protocol converter. In order to better handle the lpGBT registers, the GUI development also refers to the lpGBT control code from the lpGBT design group.

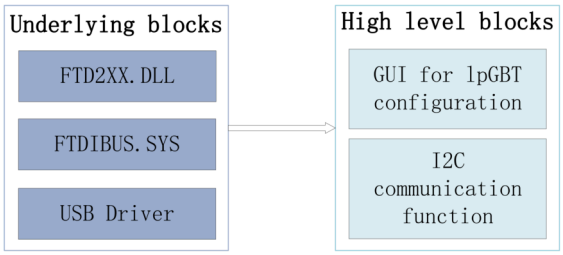


Figure 5.29: Architecture of the UPL software design.

Figure 5.30 shows a screenshot of the GUI for setting the lpGBT register values. As can be seen, each lpGBT register has a corresponding row in the GUI where both read and write operations of the register can be realized by just simply clicking on the buttons, and the operations on registers can be done either individually or collectively. In terms of register writing, apart from manually setting their values, the GUI can also support the value assignment by loading a pre-prepared configuration file.

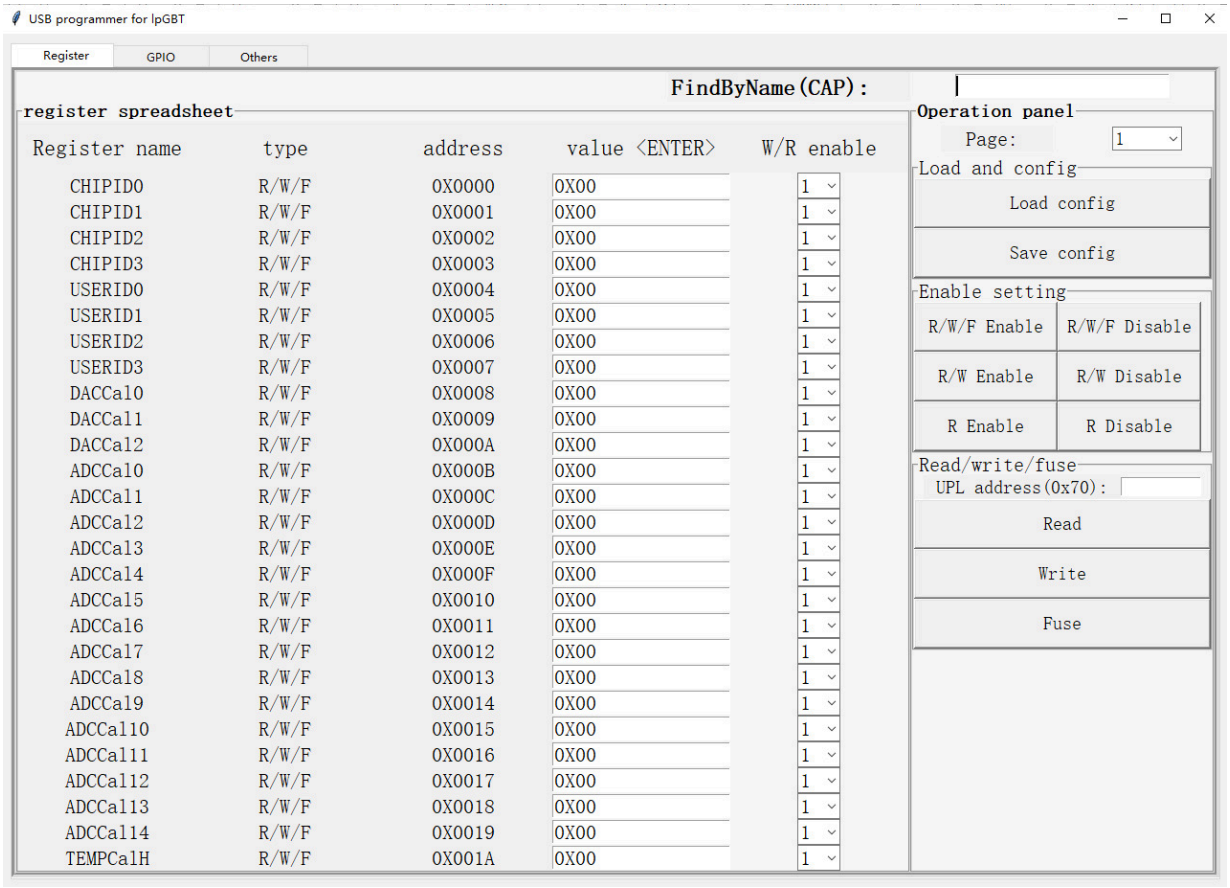


Figure 5.30: GUI for setting the lpGBT register values.

In addition to the register configuration, other kinds of configurations can also be supported by the GUI. Figure 5.31 shows a screenshot of the GUI for the lpGBT mode setting, reset and monitoring.

It's worth mentioning that equivalent python scripts are also developed to realize the same functions with the GUI in case users prefer to use computer terminal.

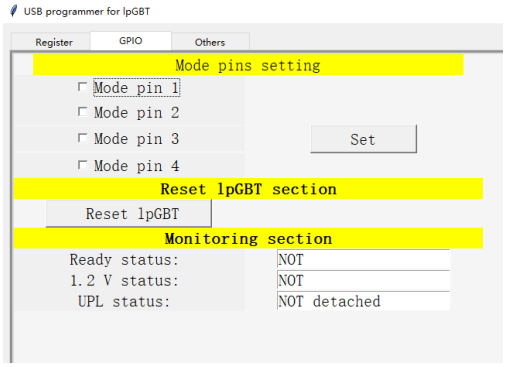


Figure 5.31: GUI for other lpGBT settings.

5.4 Firmware design

At the stage of either individual test or joint test, the firmware implementations on the FPGAs of different boards are indispensable. For example, for the individual test of the lpGBT daughter board, a FPGA-based board is planned to be used as a carrier board to hold and communicate with the daughter board under test, therefore the dedicated firmware has to be developed and implemented on the FPGA to decode/encode the data streams from/to the lpGBT; For the individual tests of the VTRx+ and SFP+ daughter boards, the Xilinx evaluation board or other custom-made FPGA-based boards can be used, it's also required that the dedicated firmware for the FPGA implementation should contain two parts, one is used to emulate the lpGBT functionality, the other works as the back-end counterpart of lpGBT to interact with lpGBT logic, such decoding and encoding the data streams. For the joint test of modular PEB system, as mentioned in Section 5.2.1, an existing MicroTCA fast control board will be used as the DAQ board of the system, hence the firmware that can serve as the back-end counterpart of lpGBT has also to be implemented on the FPGA of the DAQ board. In case of using the substitute board of the lpGBT daughter board, the firmware that can realize lpGBT functions is needed. In case of using module emulators, specific firmware has also to be developed for its FPGA to mimic the digital functions of the ALTIROC, and corresponding firmware that needs to be deployed on the DAQ side to decode/encode the data from/to module emulator, is demanded as well.

Actually, all the above-mentioned firmware developments in different scenarios boil down to building up four types of firmwares. The first one is the firmware for the module emulator,

aiming to mimic the digital function of ALTIROC. The second one is the firmware that works as the back-end (or DAQ) counterpart of ALTIROC function, aiming to decode the ALTIROC timing data stream. The third one is the firmware that can realize the function of lpGBT. The fourth one is the firmware that works as the back-end (or DAQ) counterpart of lpGBT, aiming to decode the lpGBT uplink data stream and encode its downlink data stream. The first and second ones are presented in Section 5.4.1 while the third one fourth one are respectively presented in Section 5.4.2 and Section 5.4.3.

5.4.1 Firmware design for module emulator and its DAQ counterpart

5.4.1.1 Firmware design for module emulator

The firmware for module emulator aims to imitate the ALTIROC functions. The idea is to develop the firmware based on the ALTIROC HDL (hardware description language) code from the ASIC design group. According to the test requirements of different stages, the logics of different ALTIROC functions can be integrated into the firmware, such as the timing data path, luminosity data path, slow control and fast command. Of course, the most important and basic function is to transmit timing data of particle hits, which was firstly implemented into the module emulator at the early stage of the modular PEB test. Figure 5.32 shows the block diagram of the logic handling the ALTIROC timing data.

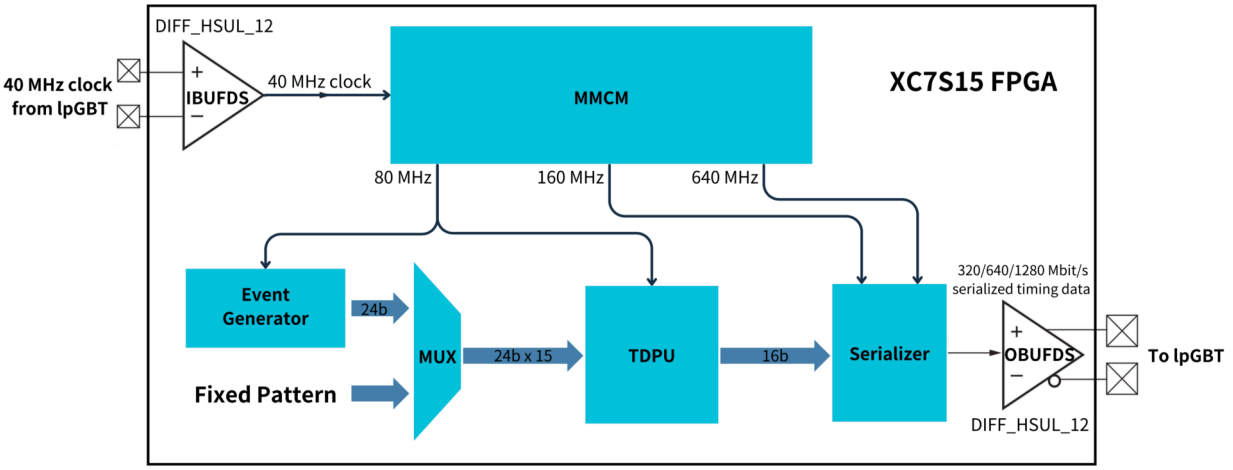


Figure 5.32: Block diagram of the logic handling the ALTIROC timing data.

As shown in the block diagram, a MMCM IP core is instantiated to regulate the 40 MHz input clock from lpGBT to three different clocks (80 MHz, 160 MHz and 640 MHz) for the other logics. The input differential clock is firstly processed by the input buffer IBUFDS with the I/O standard DIFF_HSUL_12 and becomes a single-ended clock as the input of the MMCM. There are two sources of timing data, one is from the dedicated event generator while the other one can be a fixed data pattern. One of them is selected and goes to the

TDPU (trigger data processing unit) as input. Since the TDPU requires fifteen inputs representing fifteen columns of the ALTIROC pixel matrix, the one input data is replicated by fifteen times to feed into the fifteen inputs, which will be wrapped into the special header-events-trailer data frame by circularly triggering the columns. Then the formatted data will be encoded and sent out to serializer, which can serialize the data with the width of sixteen bits to the data stream with the speed of 320 $Mbit/s$, 640 $Mbit/s$ and 1280 $Mbit/s$ and sent it to the output buffer OBUFDS. Finally the OBUFDS converts the single-ended data stream into differential data stream that goes to lpGBT at the end.

The block TDPU is mainly developed by transplanting the HDL code from ALTIROC design group. Figure 5.33 shows a block diagram of the TDPU logic. It is consisted of two parts, the first part is the hitDataFormatting logic used for packaging data frame in the format of header-events-trailer (or SOF-events-EOF), where the SOF (start of frame) records the data type (timing data or test pattern) that will be filled into the frame, the bunch crossing ID and trigger ID while the EOF (end of frame) records the number of events that are already filled into the frame, and the 8-bit CRC (cyclic redundancy check) calculated from the 24-bit input data event that was lastly filled into the frame. At the end of the logic, the frame is send out with the data width of 16 bits under the clock of 80 MHz. The other part of the TDPU is the dataEncode logic, which firstly splits the 16-bit data from last logic into two bytes. Then each of them is processed by a 8b10b encoder. The gear box is followed to adjust the bandwidth difference before and after the encoding, and it will send out the data again with the width of 16 bits under the clock of 80 MHz and feed back a ready signal to the previous logic for the acknowledgement of its availability of processing new data stream. At the end of the logic, there is a 2-to-1 MUX for deciding if the original data or encoded data is sent out to the next block, and its strobe signal is the enable pulse of the 8b10b encoding.

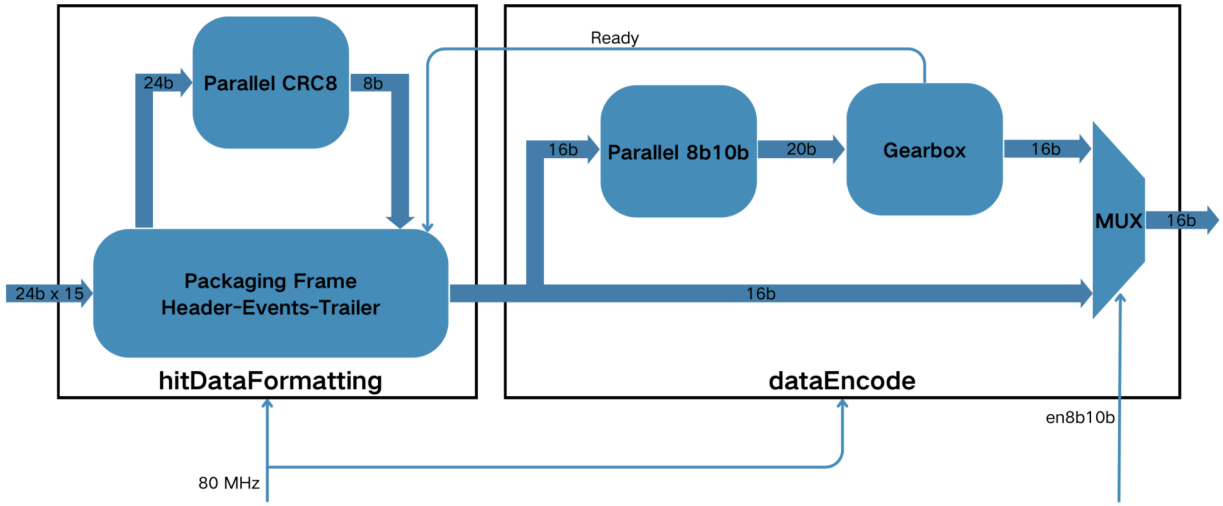


Figure 5.33: Block diagram of the TDPU logic.

In terms of the event generator, there are also two configurable ways of generating random timing data, one is to use the PRBS logic to generate pseudo data while the other is to circularly load the value entries stored in a block RAM (random access memory) of the FPGA. The block RAM is actually instantiated with the help of an IP core called block memory generator, where many parameters of the memory can be customized, such as the number of ports for memory access, the width and depth of the memory, and the data file used for initializing the memory. In terms of the last block of the timing data processing, the serializer will firstly sample the incoming data according to the transmission rate setting then get serialized by the Xilinx I/O resource OSERDESE2, which requires the high-speed clock of 640 MHz and the divided clock of 160 MHz.

5.4.1.2 Firmware design for DAQ counterpart

The corresponding ALTIROC firmware on the DAQ side aims to decode the data stream from the front-end module emulator. Is is designed to implement the inverse process of the treatment that the data gets at the frontend by the ALTIROC logics. It should be used in the FPGA together with other DAQ logics in the backend of the modular PEB system. The main clock of the logic is 160 MHz, which benefits the handling of the data in byte mode, such as the 8b10b decoded data and the 8-bit CRC data. The clocks required by the logic are provided by the clock distribution solution on the DAQ side.

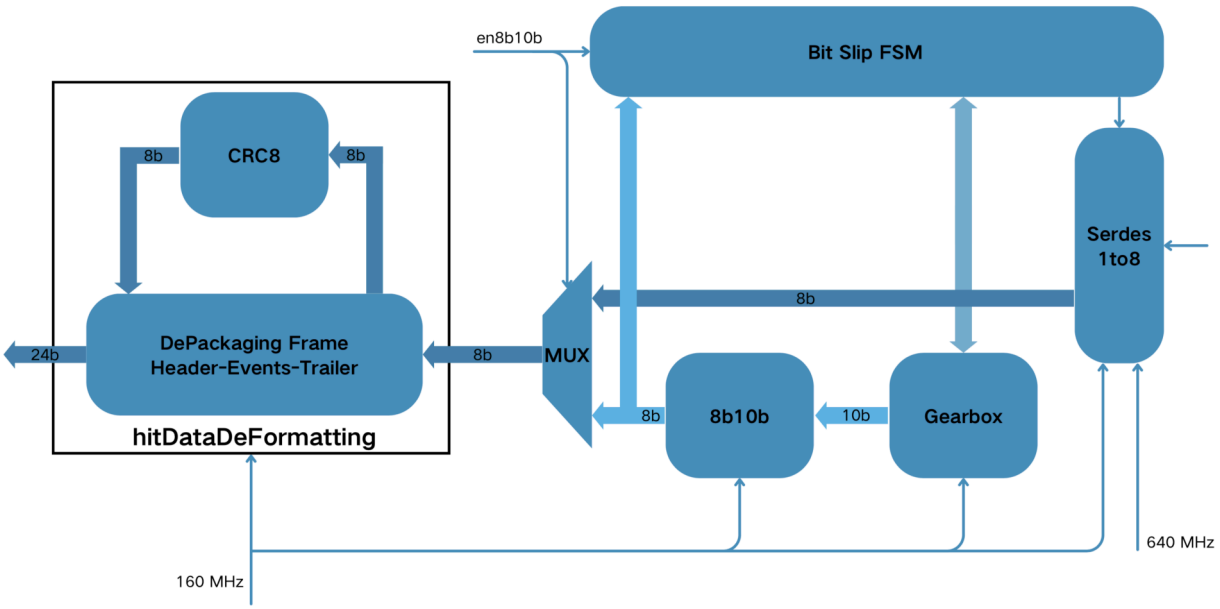


Figure 5.34: Block diagram of the DAQ counterpart logic of the module emulator.

Figure 5.34 shows the block diagram of the DAQ counterpart logic of the module emulator. The data flows from right side to left side. It can be divided into two parts, the first part, hitDataDecoding logic, is on the right side while the other part, hitDataDeFormatting logic, is on the left side of the block diagram. The serialized data stream is firstly processed

by the deserializer into the data with the width of eight bits. In the case of using the 8b10b encoding on the module emulator, the deserialized data byte is fed into the gearbox that can convert the byte into 10-bit width data, which then goes to the 8b10 decoder whose output is data byte again. Then the decoded byte as the input of the bit slip FSM, which can detect if the current byte delimitation of the deserializer is correct or not and keep feeding bit slip signal back to the deserializer until the correct byte delimitation is detected. Finally the decoded byte is sent to the next logic. In the case of not using the 8b10b encoding, the deserialized data byte directly goes to the bit slip FSM, which will do the same operation with previous case until the correct byte delimitation is detected. At the end the deserialized data byte is also sent to the next logic hitDataDeFormatting. In the hitDataDeFormatting logic, the header-events-trailer data frame is recovered and the 8-bit CRC is compared with the newly-calculated CRC based on the received event. Finally, if the CRC check is fine, the timing data events will be sent out.

5.4.2 Firmware design for lpGBT emulator

5.4.2.1 Overview

The firmware for lpGBT emulator aims to realize the digital functions of the lpGBT, especially the uplink and downlink datapaths. It is developed based on the HDL code from the lpGBT design group. In the case of lpGBT shortage or some special testing scenarios, the firmware can be implemented in the FPGA of the lpGBT emulator, such as the substitute of the lpGBT daughter board or Xilinx evaluation board.

Figure 5.35 shows an overview of the firmware for lpGBT emulator. The firmware can be generally divided into two parts, the first part on the upper side is used for handling the downlink data stream while the second part on the lower side is used for handling the downlink data stream. The emulated lpGBT should work in the transceiver mode, and support $10.24\ Gbit/s$ uplink bandwidth and FEC5 encoding strategy. The high-speed data streams are handled by the Gigabit transceiver (GTX) embedded in the FPGA, whose uplink and downlink data rates should be the same and set to be $10.24\ Gbit/s$. Since the downlink data stream of lpGBT is just $2.56\ Gbit/s$, it is quadrupled to be able to match the GTX downlink data rate.

In terms of the uplink data path, starting from the left side of the block diagram, just like the real lpGBT there are seven E-ports followed by seven SerDes (serializer/deserializer). Since each E-port has the bandwidth of $1.28\ Gbit/s$, the SerDes sends out the 32-bit data under the drive of the $40\ MHz$ clock. Then the data goes to the 32-bit buffer register. Together with the 2-bit EC data, the seven 32-bit data are packaged into an intermediate 234-bit data frame containing 2-bit IC, 2-bit EC, 6-bit LM and 224-bit data. After further processed by the uplink logic, the final 256-bit data frame (including 2-bit header, 234-bit

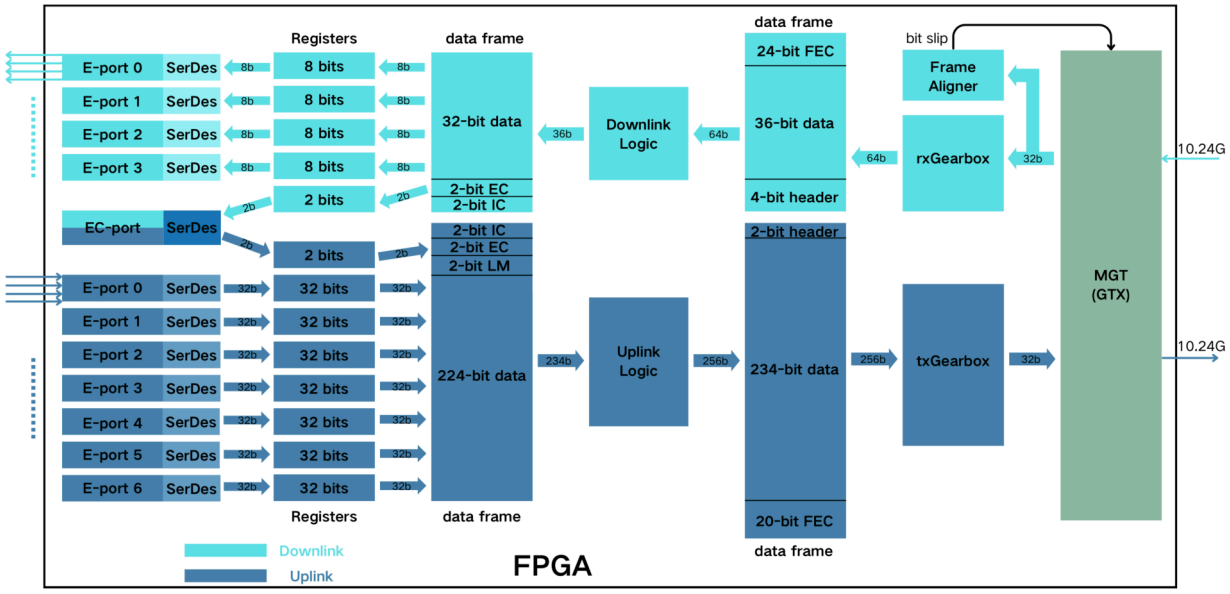


Figure 5.35: Block diagram of the firmware for lpGBT emulator.

data and 20-bit FEC) are sent out to the transmitter gearbox, which can convert the 256-bit data frame at the clock of 40 *MHz* to the 32-bit data segment at the clock of 320 *MHz*. Finally the 32-bit data is fed into the GTX and serialized into the data stream with the speed of 10.24 *Gbit/s*.

In terms of the downlink data path, starting from the right side of the block diagram, the serialized downlink high-speed data stream (10.24 *Gbit/s*) with just 2.56 *Gbit/s* valid information embedded, is firstly processed by the GTX and deserialized into the 32-bit data that is driven out by the 320 *MHz* clock. Then the data goes to both the frame aligner block and receiver gearbox block. The frame aligner is used to find the header of the frame by feeding the bit slip pulse back to the GTX until the data frame is aligned. After being sampled by the receiver gearbox every four 320 *MHz* clock cycles, the valid information is extracted out and further sent out in the form of the 64-bit data frame, including 24-bit FEC, 36-bit data and 4-bit header. Then the data frame is processed by the downlink logic and an 36-bit intermediate data frame (including 32-bit data, 2-bit EC and 2-bit IC) is popped out. The 2-bit EC is deserialized and goes to the EC-port while the 32-bit data is evenly split into four pieces, which are deserialized and go to the four E-ports. About the exact data rate of each E-link, it is completely dependent on the data rate settings of the E-ports.

5.4.2.2 Uplink and downlink logics

The uplink and downlink logics are the key parts of processing the uplink and downlink data streams. The main clock of them is the 40 *MHz* clock. Figure 5.36 shows the more detailed block diagrams of the two logics.

The uplink logic is consisted of three main blocks, the scrambler, FEC5 encoder and

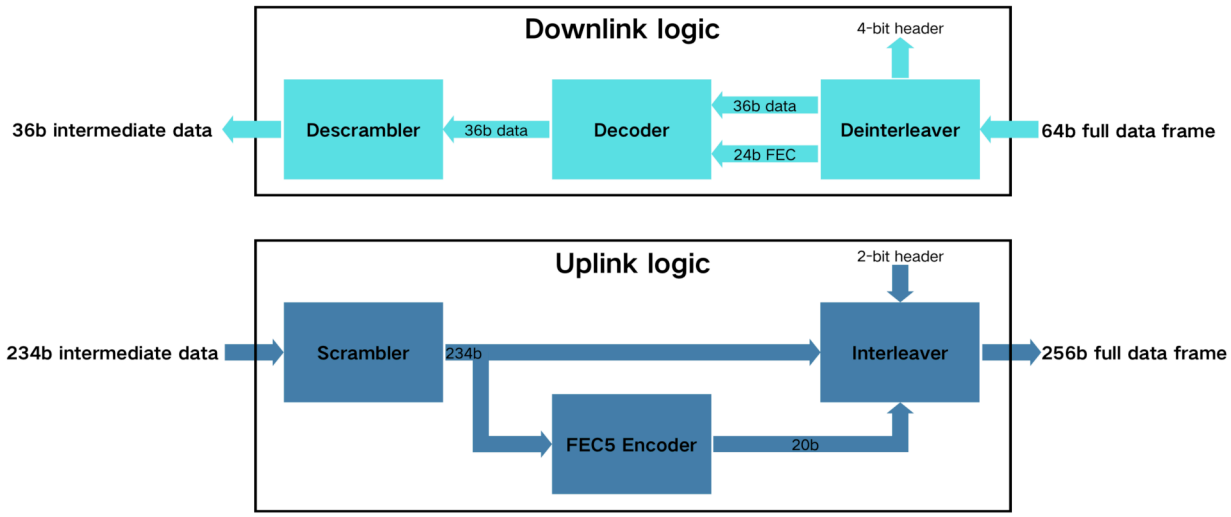


Figure 5.36: Block diagram of the uplink and downlink logics.

interleaver. The 234-bit intermediate data frame is firstly pushed into the scrambler and transposed in a certain way, and the output from the scrambler is a scrambled 234-bit data frame, which goes to the FEC5 encoder for the computation of a 20-bit error correction code using the FEC5 method and also the interleaver. Then together with the 20-bit FEC5 code and the 2-bit header, the 234-bit data frame is interleaved by the the interleaver. Finally, the 256-bit full data stream is popped out.

The downlink logic is also consisted of three main blocks, the deinterleaver, decoder and descrambler. The 64-bit full data frame is pushed into the deinterleaver and an inverse process of the interleaving performed at the DAQ side is carried out, then the deinterleaved data frame is divided into three segments, the 4-bit header, 36-bit data and 24-bit FEC. The last two data segments are feed into the decoder, where the data is not only decoded in an inverse way of the encoding on DAQ side but also checked in correctness by comparing the newly-calculated FEC from the data to the received FEC. Then the decoded and correct 36-bit data is sent to the descrambler and descrambled in an inverse way of the scrambling performed in the DAQ side. Finally the 36-bit intermediate data frame is sent out.

5.4.2.3 High-speed interface

As shown in Figure 5.35, the high-speed interface includes GTX resource, receiver gearbox and transmitter gearbox. The GTX is the customized 7-series FPGA IP core, which can be instantiated by the dedicated transceiver wizard provided by the Xilinx software Vivado. With the help of the wizard, many key parameters can be easily adjusted. For example, the type of the transceiver can be selected from several options, such as the GTP (6.6 *Gbit/s*), GTX (12.5 *Gbit/s*), GTH (13.1 *Gbit/s*) and GTZ (28.05 *Gbit/s*). In our case, the GTX is selected since its maximum line rate can reach up to 12.5 *Gbit/s* that is adequate for the lpGBT high-speed uplink and downlink data transmission. The transceiver resource that is

used for the IP core implementation can be specified among those ones located on both the left and right side of the FPGA die, also line rates of the transmitter and receiver can be set and the clock source of this transceiver comes from the on-board clock that is specially used for the FPGA transceiver quad. For this design, the line rates of the GTX transmitter and receiver are both set to be 10.24 $Gbit/s$ while the external data widths of them are set to be 32 bits.

The receiver gearbox receives the 32-bit external data popped out of the GTX under the clock of 320 MHz, and samples the data also with a 320 MHz clock. A 256-bit register will be filled up every eight sample clock cycles and the 64-bit valid data is directly extracted out of the register by evenly taking a quarter of the bits. Then the 64-bit full data frame is sent to the downlink logic at the clock of 40 MHz. Of course, the correct data delimitation is determined by the frame aligner. The mechanism of the frame aligner is to keep generating bit slip pulse to GTX until the correct delimitation is detected. In the frame aligner, there are eight 32-bit buffer registers in a row used as the sample data storage. It checks every bit of the first register, and generates bit slip pulse if the correct bit is not found. Then it will do the same operation to other seven registers until the header of the data frame is found. After the frame alignment, the frame aligner will keep monitoring the data stream, once the misalignment is detected, it will start to work until a new alignment is achieved.

The transmitter gearbox receives the 256-bit full data frame from the uplink logic and samples them with the clock of 40 MHz, then the frame is evenly divided into eight 32-bit data segments that are driven out of the gearbox with the clock of 320 MHz. All the eight data segments can be sent out by eight 320 MHz clock cycles, which is exactly the duration of one 40 MHz clock cycle. Finally the 32-bit data segment is fed into the GTX as its external data input and serialized into the data stream of 10.24 $Gbit/s$.

5.4.3 Firmware design for lpGBT DAQ counterpart

5.4.3.1 Overview

The firmware of the lpGBT DAQ counterpart aims to implement the inverse processes of the lpGBT logics, especially processing the downlink data stream before sending it to the lpGBT downlink logic and the uplink data stream from the lpGBT uplink logic. It is also developed based on the HDL code from the lpGBT design group. The firmware is expected to be used on the FPGA of DAQ board, such as the MicroTCA fast control board chosen as the DAQ board of the modular PEB system, and Xilinx evaluation board used for testing lpGBT daughter board and other types of daughter boards.

Figure 5.37 shows an overview of the firmware for lpGBT DAQ counterpart. Generally speaking, the firmware can be split into two main parts, the first part on the upper side is used to pre-process the downlink configuration data before it's distributed to the lpGBT

logic while the other part on the lower side is used to recover the timing and luminosity data frames that was built by the lpGBT logic. The high-speed data streams are also handled by the FPGA Gigabit transceiver (GTX), whose line rates of transmitter and receiver are all set to be 10.24 *Gbit/s*. Since the configuration downlink bandwidth required by the detector system is just 2.56 *Gbit/s*, it is quadrupled to be able to match the GTX transmitter line rate. The quadruple method is a good solution to the asymmetry of the uplink and downlink data rates, and its feasibility has been validated. In the case of real lpGBT ASIC that receives the quadrupled downlink, the data bits can still be correctly sampled. The data stream is actually still the required 2.56 *Gbit/s* from the perspective of the lpGBT, because every bit is repeated by four times in a row with the quadruple method. In the case of the lpGBT emulator firmware that receives the quadrupled downlink, only a quarter of the data bits are evenly sampled in its logic so the required 2.56 *Gbit/s* configuration information can also be recovered.

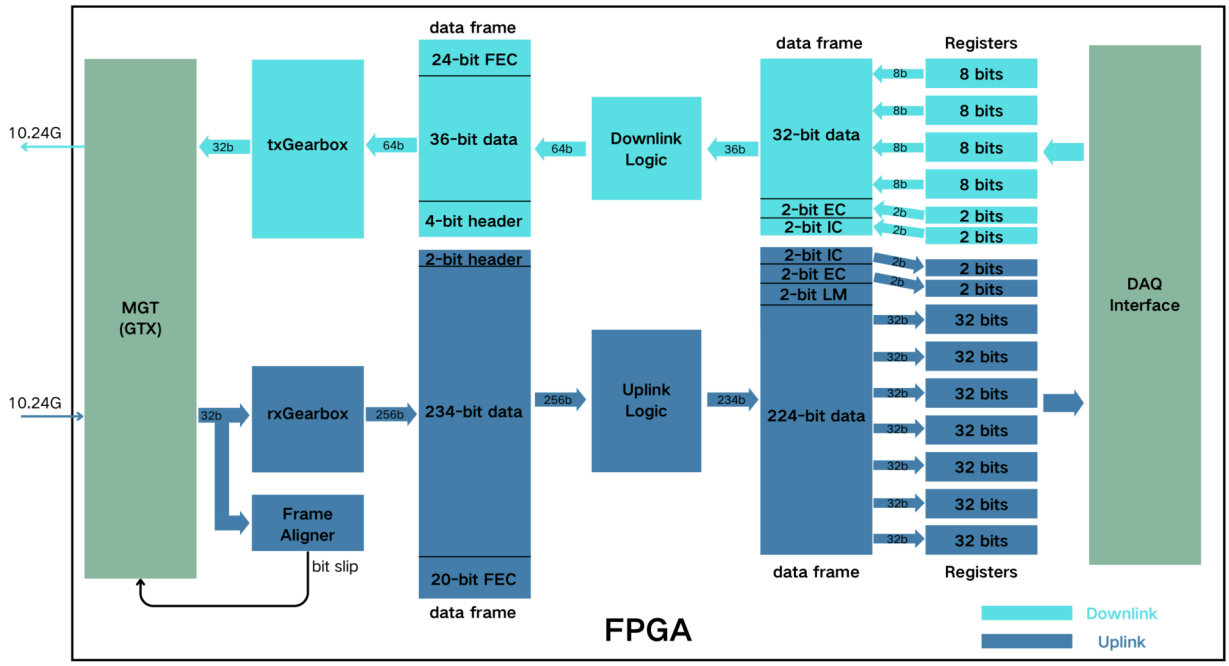


Figure 5.37: Block diagram of the firmware for lpGBT DAQ counterpart.

In terms of the uplink data path, starting from the left side of the block diagram, the serialized high-speed data stream at the speed of 10.24 *Gbit/s* goes to the GTX and gets deserialized into the 32-bit data segment, which further goes to the frame aligner and receiver gearbox. The former aims to probe the correct bit delimitation by feeding the bit slip pulse back to the GTX until the frame is correctly aligned while the latter aims to convert the data width from 32 bits to 256 bits with the bandwidth unchanged. The 256 bits are exactly one data frame of the uplink stream, which is consisted of 2-bit header, 234-bit data and 20-bit FEC. Then the entire data frame is sent to the uplink logic, whose output will be an intermediate data frame consisting of 2-bit IC, 2-bit EC, 2-bit LM and

224-bit data. Finally, those pieces of bits are pushed into their corresponding buffer registers waiting for further operation of the other DAQ logics. For example, the 224-bit data are evenly split into seven parts, each of which is temporarily stored in a 32-bit register.

In terms of the downlink data path, starting from the right side of the block diagram, other DAQ logics firstly push data bits into several buffer registers including four 8-bit registers and two 2-bit registers, then those bits are further loaded to build an intermediate 36-bit data frame, which is composed of 2-bit IC, 2-bit EC and 32-bit data. After being processed by the downlink logic, the data frame becomes a 64-bit full data frame consisting of 4-bit header, 36-bit data and 24-bit FEC, and it is fed into the transmitter gearbox where the quadruple method is implemented and a 32-bit data segment will be sent out to the GTX. Finally, the GTX serialize the data into 10.24 $Gbit/s$ data stream.

5.4.3.2 Uplink and downlink logics

The uplink and downlink logics are the key parts of handling the uplink and downlink data streams. The main clock of them is the 40 MHz clock. Figure 5.38 shows more detailed block diagrams of the two logics.

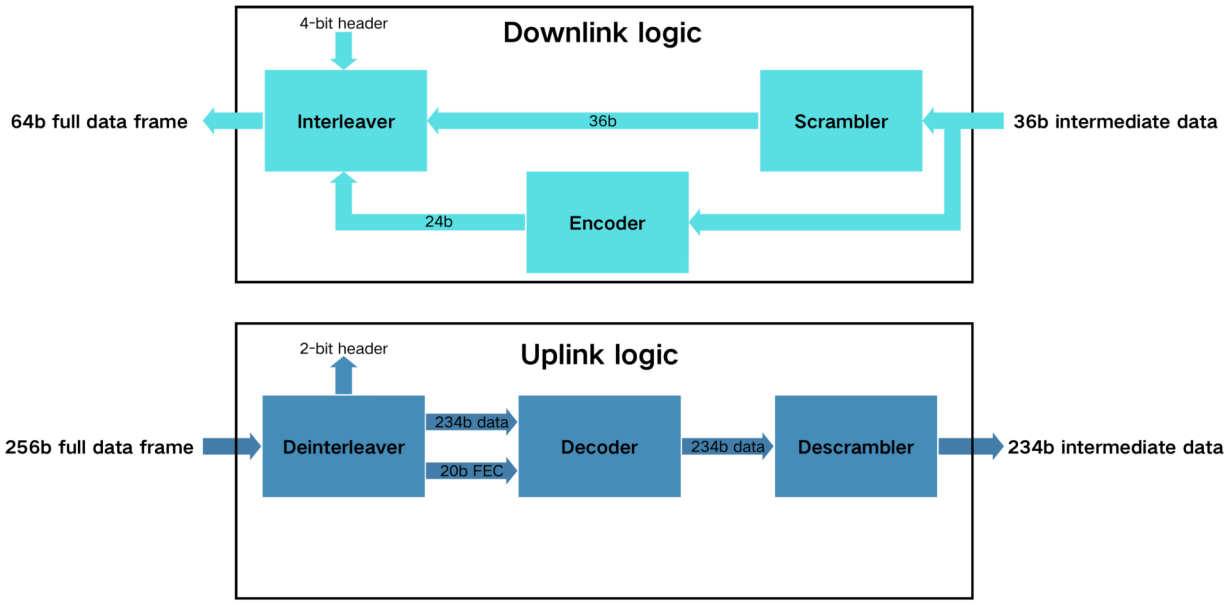


Figure 5.38: Block diagram of the uplink and downlink logics.

The uplink logic is composed of three main blocks, the deinterleaver, decoder and descrambler. The 256-bit full data frame is firstly fed into the deinterleaver, which is used to recover the data frame before the interleaving performed by the lpGBT uplink logic. After the processing of the deinterleaver, there are three data segments that are sent out, including the 2-bit header, 20-bit FEC and 234-bit data. The last two segments are further processed by the decoder, where the 234-bit data is firstly checked in correctness by comparing the newly-calculated FEC to the received 20-bit FEC and decoded to the data before the encod-

ing performed by the lpGBT uplink logic once the check is fine. Then the 234-bit decoded data goes to the descrambler and gets recovered to the original data before the scrambling of the lpGBT uplink logic. Finally, the 234-bit intermediate data frame is sent out of the uplink logic.

The downlink logic is also composed of three main blocks, the scrambler, encoder and interleaver. Starting from the right side of the downlink logic, the 36-bit intermediate data goes to both the scrambler and encoder. The former aims to transpose the bits of the data in a certain way and output the data with the same width while the latter is used to compute the 24-bit error correction code of FEC type from the 36-bit input data. Then together with the 4-bit header, the 36-bit scrambled data and the 24-bit FEC code are fed into the interleaver where all the input bits are interleaved in a special way. Finally, the 64-bit full data stream is sent out of the downlink logic.

5.4.3.3 High-speed interface

As shown in Figure 5.37, just like the high-speed interface of the firmware for lpGBT emulator, the high-speed interface of this firmware also includes GTX resource, receiver gearbox and transmitter gearbox. For the GTX resource, its characteristics are already presented in Section 5.4.2.3, so it's not described in this section.

Under the clock of 320 MHz, the receiver gearbox picks up the 32-bit external data coming from the GTX receiver and sends it to the 256-bit buffer register. Every eight 320 MHz clock cycles, the buffer register is filled up and immediately sent out with the frequency of 40 MHz. To make sure that the 256 bits that are stored in the buffer register exactly form a full data frame, the header of the data has to be correctly detected, which is carried out by the frame aligner. Just like the one used in the logic for lpGBT emulator, this frame aligner works in the exactly same way, which is to inform the GTX receiver of doing bit-slipping until the data frame header is found.

The quadruple method, that is used to solve the issue of implementing the asymmetrical uplink and downlink data rates of the lpGBT into FPGA, is used in the transmitter gearbox. The 64-bit data frame goes to the gearbox with the frequency of 40 MHz and gets sampled by the clock with the same frequency. Then 64-bit data is evenly divided into eight pieces, each of which is quadrupled in the way that each bit is repeated four times in a row. Therefore, each piece is expanded to 32 bits after the quadruple. The each 32-bit data is driven out under the clock of 320 MHz. Within one 40 MHz clock cycle, all the eight expanded pieces can exactly be sent out by eight 320 MHz clock cycles. At the end, the 32-bit data goes to the GTX transmitter as its external input data.

Chapter 6

System Tests

This chapter firstly presents the test strategy of the modular PEB, followed by the individual tests of its components. After making sure all the individual components can function well, a joint test system is set up, targeting the tests of the five key aspects, which are system configuration, data transmission, power supply, monitoring network, clock and fast command distribution. Finally, a summary is given and a prospect of how the modular PEB can be used is also presented.

6.1 Test strategy

As described in Chapter 5, the design of the modular PEB system is a large project and it is challenging and quite complicated. There are not only a large number of daughter boards that need to be designed or prepared, but also some firmwares required by the system during the test and debugging stage. It takes more than two years to finish the development of both the hardwares and firmwares and also other necessary preparations, such as the investigations of possible test strategies of the system in both the collective and individual ways, the preparations of some paramount equipments required by the future tests, and the accumulation of corresponding expertise and experience. For the test of such a complex system, it has to be carefully carried out in a systematic and progressive approach. The main idea of the test strategy is to firstly test all the pieces of the system individually including the daughter boards and the firmware implementations before the final integration of all single components, which can help us to easily pin down the sources of possible bugs. This strategy will definitely make our lives easier even through some of the individual tests seem tedious and unnecessary. At the stage of joint test, the individual pieces should be integrated into the system step by step instead of assembling the complete modular PEB system in one go.

In terms of the individual tests, the specific strategy for each of them really depends on the component itself under the test, but the general idea should be always to make sure the ability of realizing its requested functions reliably and accurately. There are plenty of

ingredients that require thorough tests before the joint test of the system. Those ingredients can be divided into two parts, the hardwares and firmwares. The former includes the lpGBT daughter board, VTRx+ daughter board, FH26W daughter board, the daughter boards of bPOL12V and TPS56428, SFP+ daughter board, the module emulator and the dedicated lpGBT programmer UPL. The latter includes several types of FPGA logics, which are respectively for the module emulator, lpGBT emulator, and their DAQ counterparts. Since some of those ingredients are related to each other, their tests can be performed with just one experiment setup, which can significantly reduce the workload of the hefty individual tests. For example, in the setup used for VTRx+ daughter board test, the SFP+ daughter board, together with the firmwares for both the lpGBT emulator and its DAQ counterpart, can also be tested efficiently. The individual tests are presented in Section 6.2.

In terms of the joint test, the idea is to assemble all the components together and check the performances of many different key parts with the monolithic modular PEB system. For example, those points raised up in the requirement analysis Section 5.1 can have final conclusions or crucial enlightenments that can guide the following plan of the PEB research and development. Along the test of the modular PEB system, there are different versions of system setups, which are generally evolving from being simple to becoming fully-fledged. At the very beginning of the integration, just one single module emulator is attached to the carrier board and the SFP+ daughter board is used to mimic the VTRx+ daughter board. Gradually, an increasing number of components are integrated into the setup, and the system becomes more and more complete. After each intermediate setup, some more tests are performed. In such a way, intractable issues can be easily spotted and solved when the system is still simple, and we gain valuable experiences and better understandings about the system along the journey from simple to complex. The joint test is performed based on the complete system, but it's worth mentioning that the successful operation of the complete system is not accomplished overnight. There are also plenty of items that need to be tested during the joint test, such as system configuration, data transmission, power supply distribution, monitoring, clock and fast command distribution, and other kind of performance tests.

The MicroTCA fast control board (uFC) [162] is largely used as the DAQ board in both the individual tests and joint test. It is a Xilinx Kintex-7 FPGA-based system with a lot of expansion interfaces for diverse requirements of users. The uFC can provide users with the data memory, trigger, reference clocks and SFP+ connectors that are commonly-used in general experiments. Most importantly, it has two FMC connectors each with a large array of configurable I/Os and high-speed links with the data rates of up to $10\ Gbit/s$, therefore most of our daughter boards can be tested with the help of this board. For example, the firmwares for lpGBT emulator and its DAQ counterpart can be implemented in the FPGA while the VTRx+ or SFP+ daughter boards are connected to the two FMC connector,

so that all of them can be tested with uFC. Of course, there are more tests that can be performed by this board, and they will be presented in the following sections. Figure 6.1 shows a picture of the board where its two FMC connectors on the right side are occupied by two lpGBT daughter boards, the FPGA in the middle area is equipped with cooling fin for a proper heat dissipation caused by the possibly-large amount of implemented logics, and also the two SFP+ connectors on the upper-right corner can be attached by Ethernet cables or optical fibers to interact with the host computer for future data storage and further analysis.

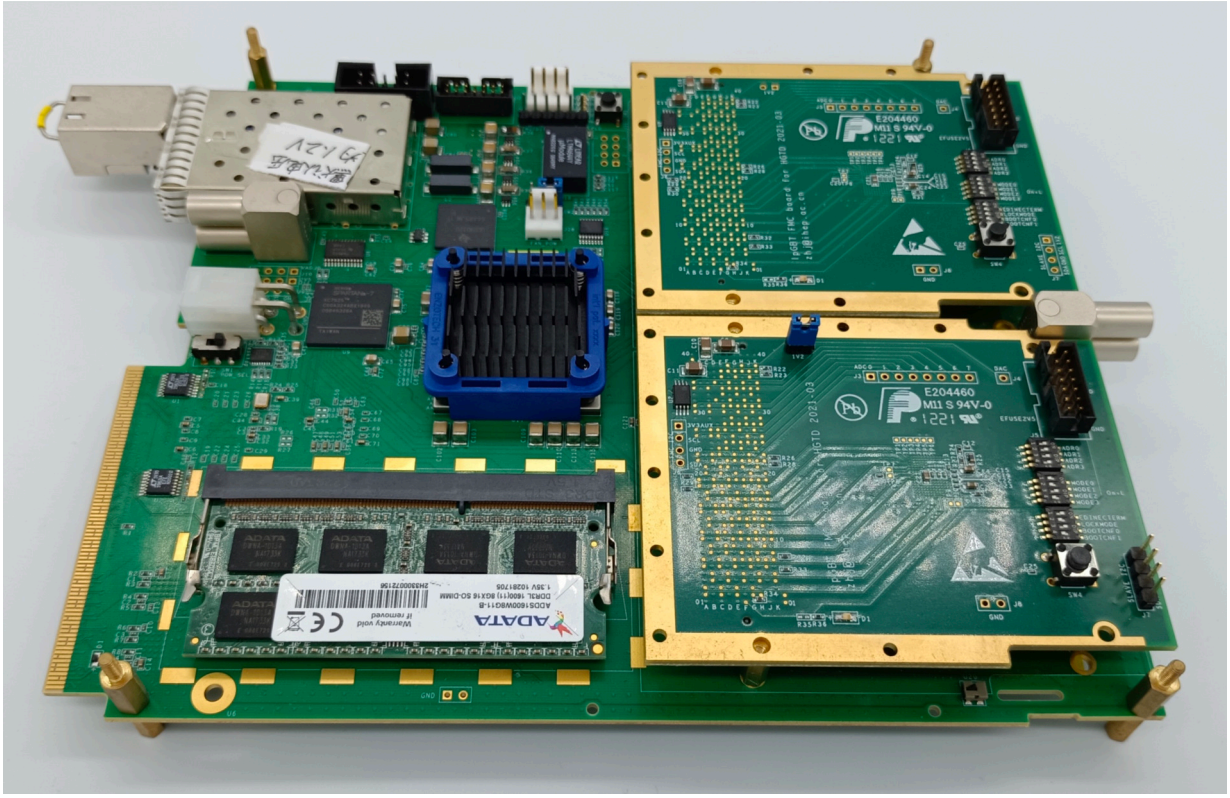


Figure 6.1: Picture of the uFC with its two FMC slots occupied by two lpGBT daughter boards.

6.2 Individual test

6.2.1 Test of hardware components

This section describes the tests of the main hardware components of the modular PEB system, including the lpGBT daughter board, VTRx+ daughter board, bPOL12V daughter board. The test of other components can be found in Appendix B.

6.2.1.1 lpGBT daughter board test

The lpGBT daughter board holds only one main ASIC, which is the lpGBT. The test of the board can be divided in to two parts, the lpGBT configuration test and the FMC connectivity test. The former is the prerequisite of all the other operations to this board since the lpGBT is operational only when it's properly configured. The latter can make sure that the lpGBT daughter board has the complete and reliable connections to the carrier board. The important signals during the FMC connectivity test include the 28 uplink E-links, 16 downlink E-links, one high-speed uplink and one high-speed downlink.

Figure 6.2 shows the setup for lpGBT daughter board test. The uFC board is used as the carrier board of the lpGBT daughter board. There are two lpGBT daughter boards that are inserted into the uFC, but only the upper one is tested. The uFC board is powered by the power supply unit that has a $12\ V$ voltage output. Since the lpGBT ASIC is connected, the firmware implemented in the FPGA of uFC should include the logics for the lpGBT DAQ counterpart to be able to handle the communication with the real lpGBT logic. Therefore, the complexity of the firmware is quite large, which leads to a sizable FPGA resource utilization ratio thus a hefty power consumption of the FPGA. A cooling fan is used towards the FPGA to accelerate the heat dissipation of it. The firmware from the host computer is programmed into the FPGA through the JTAG communication, which is carried out by the USB cable with one end connected to the host computer and another end connected to a Xilinx JTAG programmer that is further attached to the JTAG connector on the uFC board. The lpGBT configuration is performed by the UPL that is connected to the lpGBT daughter board through the flat cable. On its other side, the UPL is further connected to the host computer.

Figure 6.3 shows the architecture of the firmware developed for the uFC FPGA. As can be seen, a large portion of the FPGA logics is the firmware for the lpGBT DAQ counterpart, which is already described in Section 5.4.3. Generally speaking, together the lpGBT logic, the FPGA logics handle two loop paths, the uplink loop path and downlink loop path. For the uplink loop path, starting from the upper side of the block diagram, there are 28 FPGA OSERDES2 resources that are used to serialize 28 8-bit fixed patterns into serialized data streams with the speed of $320\ Mbit/s$, $640\ Mbit/s$ or $1.28\ Gbit/s$ and send them to the eight lpGBT uplink E-ports, where its 28 E-links are exactly connected to the 28 serialized data streams. After the processing of the lpGBT uplink logics, all of those input data are wrapped into data frames and finally serialized into a $10.24\ Gbit/s$ high-speed data stream that is sent back to the FPGA. Inside the FPGA, the lpGBT DAQ counterpart logic handles the input high-speed data stream and finally recover it to the original fixed patterns stored in those 32-bit buffer registers, which can be easily checked with ILA logic and displayed on the host computer. For the downlink loop path, starting from the lower side of the block diagram, a 36-bit fixed pattern is as the input of the lpGBT DAQ counterpart logic, which first splits

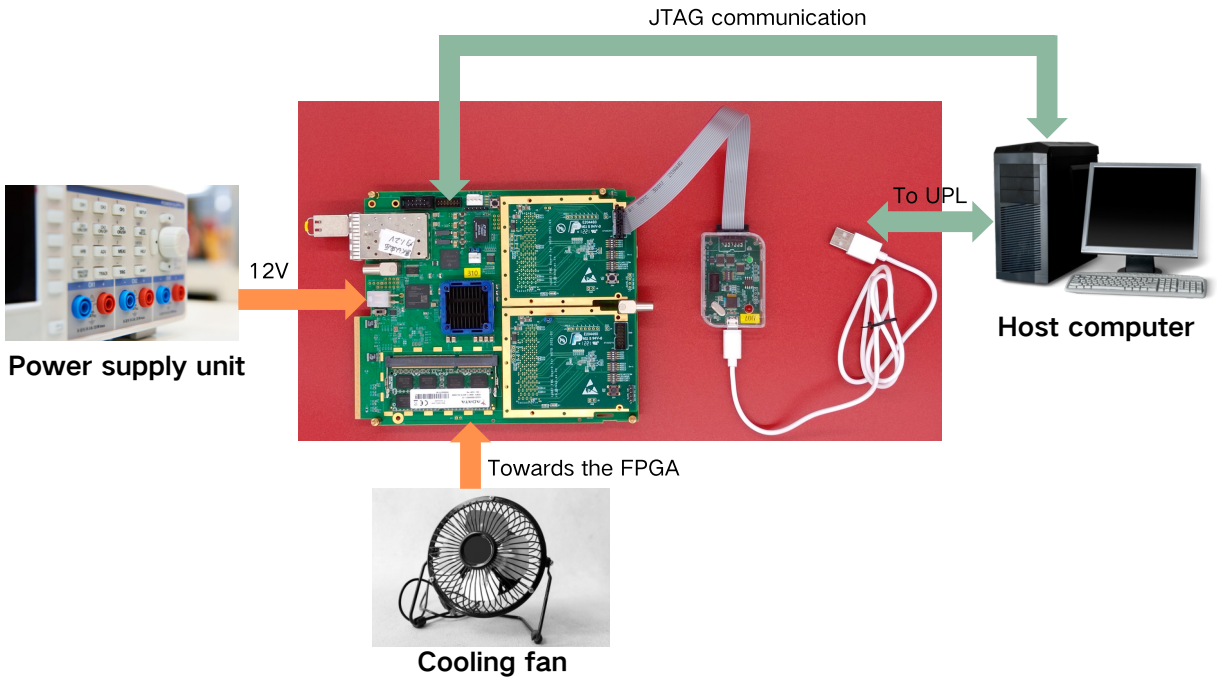


Figure 6.2: Picture of the setup for the lpGBT daughter board test.

the pattern into segments and temporarily stores them in the buffer registers. After the processing of the downlink logic, the data will be sent out of the FPGA in the form of 10.24 *Gbit/s* serialized data stream with just 2.56 *Gbit/s* valid information contained, and go to the lpGBT. Then the lpGBT downlink logic will recover it to the original fixed patterns, which are further serialized into 16 downlink E-links with the data rate of 320 *Mbit/s* by the four lpGBT downlink E-ports. The 16 E-links are connected to exact 16 FPGA ISERDES2 resources that are used to deserialize the input data streams into 8-bit fixed patterns, which can also be easily checked with ILA logic and displayed on the host computer.

To be able to handle both the uplink and downlink data paths, the lpGBT should be configured as a transceiver. The baseline encoding strategy and high-speed uplink rate for the transceiver lpGBT used for the HGTD is the FEC5 and 10.24 *Gbit/s*. Therefore, the four mode pins MODE[3:0] are set to be 4'b1011 through the dedicated dip switch on the lpGBT daughter board. The lpGBT initialization and configuration after power on are controlled by a FSM (finite state machine) logic called power-up state machine (PUSM). Only when the PUSM goes through all the states, the lpGBT is ready and operational. It jumps to the next state only if the specific condition is satisfied. Only the power supply of the ASIC is stable and larger than 0.9 *V*, it goes to the following states for configuration information loading. There are two pins (BOOTCNF0 and BOOTCNF1) that determines where the configuration information is loaded from, for example, it can come from the E-fuse array, ROM memory or register array. Another dedicated dip switch on the lpGBT daughter board can set the two pins, whose values are configured to be 2b'11 in our case to be able to load the configuration information from the lpGBT register array. Then the PUSM will stay at

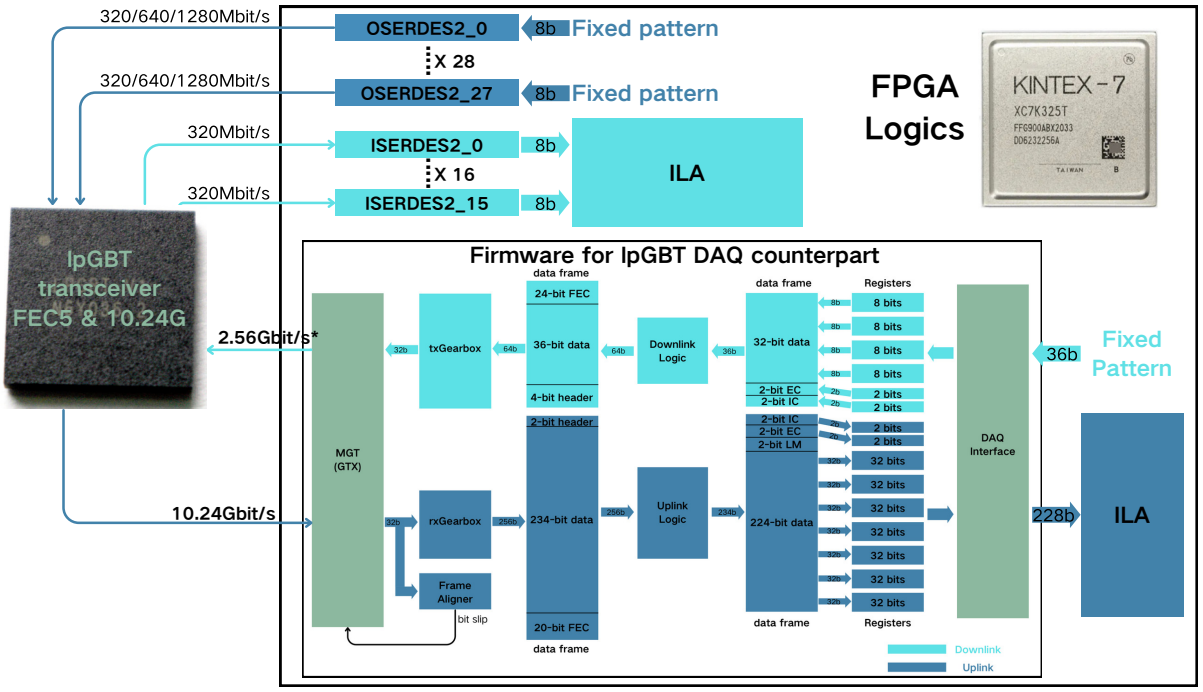


Figure 6.3: Architecture of the firmware developed for the uFC FPGA. 2.56 *Gbit/s** : the data stream is running at 10.24 *Gbit/s* but only 2.56 *Gbit/s* valid information is embedded.

the state called PauseForPllConfig until the lpGBT register configuration is completed by flipping the bits dllConfigDone and pllConfigDone of the register POWERUP2 to high as its end. With the same dip switch, the LOCKMODE pin is set to be 1'b1 to recover the primary clock from the downlink high-speed data stream. There is also another dip switch connected to four address pins ADDR[3:0] that can be used to set the four LSBs of lpGBT chip address used in I^2C or serial control while the three MSBs are set by the register called ChipAddressBar.

In terms of the lpGBT register configuration, not all the registers are required to be configured. In general, there are three types of registers, including read/write/E-fuse registers, read/write registers and read-only registers. The first type of registers can not only be read and written but also initiated by the E-fuse array, and the second type of registers can be read and written while the last type of registers can only be read. For the register configuration, only those important registers from the first and second types need to be written with proper values based on the experiment requirements. For example, the registers about the clock generator have to be set carefully since the clocks of the entire ASIC are originated from it. In this setup, all the uplink E-ports are used, so the data rates, activated E-links, termination strategies and DLL (delay locked loop) parameters of them should be configured by writing the corresponding registers. Similarly, since all the downlink E-ports are used, the data rates and activated E-links should also be configured in the same way. The registers about the line driver of the high-speed transmitter are set since they can adjust the pre-emphasis parameters which significantly determine the transmission quality of the

high-speed signal. The registers about the equalizer and frame aligner of the high-speed receiver are also set since they are of great importance on the signal reception efficiency. In addition, the registers about the 29 E-clocks and the phase shifter clocks can also be set. Finally, the two bits dllConfigDone and pllConfigDone of the register POWERUP2 have to be set to high as the end of the lpGBT register configuration.

Right after detecting the high level of the pllConfigDone, the PUSM continues going forward to do the PLL initialization. When then PLL is locked, it will go through several states for the execution of the I^2C master transaction, which aims to configure a laser driver chip or any other component in the system. After this, it directly goes to the state called PauseForDllConfig. Since the bit dllConfigDone is already set to be high, it will go through the states for DLL initialization. When the DLL is locked, it will check if the uplink and downlink E-links are locked or not. If locked, it goes to the last state called READY indicating the readiness of the lpGBT. Otherwise, it will go back to the first state to repeat the same procedure. It can also be configured that the PUSM goes directly to the READY state without checking the lock status of the uplink and downlink E-links. Finally, the READY signal pulse is sent out through the lpGBT READY pin, which is routed to the 14-pin header connector on the lpGBT daughter board. Thus, the signal pulse goes to the READY pin of the UPL and the status can be displayed on the GUI, at the meantime, the green LED of UPL gives out light as a more intuitive display. In this setup, the green LED can always be lit up after the lpGBT register configuration with the use of the UPL toolkit.

In terms of the FMC connectivity test, the idea is to check all the connectivity of the lpGBT pins that are routed to the FMC connector. With the setup and the above firmware, the pins of high-speed links, uplink and downlink E-ports can be easily checked by simply sending fixed patterns from FPGA side to the lpGBT then checking the the loop-back data again on the FPGA side. The correctness of the data that are looped back from lpGBT indicates the situation of the pin connectivity. For the connectivity test of other pins, different strategies are used but still based on the same setup and firmware implementation. For example, about the pins of E-clocks, the activated channels and their frequencies are firstly configured, then probe of the oscilloscope is attached to the corresponding test points of those pins right on the back side of the lpGBT daughter board. As shown in Figure 6.2, when the lpGBT daughter board is inserted into the uFC, its back side faces upwards and the test points can be easily accessed by the probe. From the measure waveform, the connectivity of those pins can be deduced.

6.2.1.2 Test of VTRx+ daughter board

The VTRx+ daughter board is mounted with the VTRx+ optical transceiver, which has four uplinks and one downlinks. The test of the VTRx+ daughter board is mainly to evaluate the transmission performance those high-speed links. To do that, the BERT (bit error rate

test) will be performed for those links.

Figure 6.4 shows the setup of the BERT test for the VTRx+ high-speed links. As can be seen, again the uFC board is used as the carrier board holding one SFP+ daughter board and one VTRx+ daughter board. The pig tail fiber bunch of the VTRx+ module is terminated with a MT connector, which can be mated with the MPO connector. Therefore, the pig tail is connected to the MPO-to-LC adapter, whose other side is inserted by five optical fibers with the LC connectors at the corresponding slots. Then five optical fibers are connected to the four SFP+ optical transceivers that are mounted on the SFP+ daughter board. Among the five fibers, three of them go to the receiver side of the three SFP+ optical transceivers on the right while another two of them go to the first SFP+ optical transceiver. In addition, the uFC board is supplied by a 12 V input voltage through the write power supply connector, and also its JTAG connector is occupied by a Xilinx JTAG programmer for downloading the firmware into its FPGA.

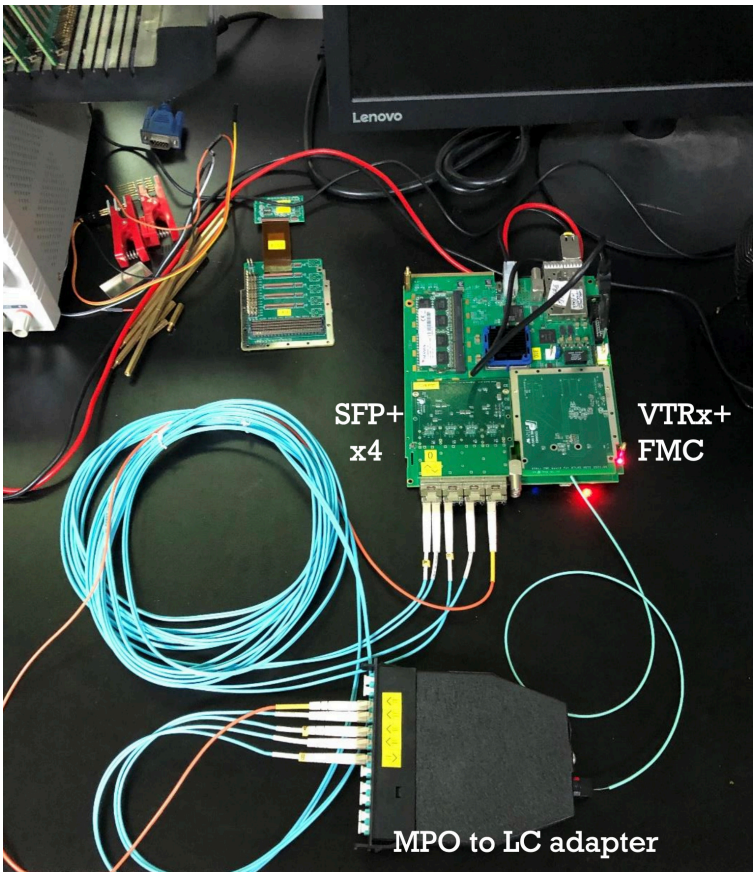


Figure 6.4: The setup of the BERT test for the VTRx+ high-speed links.

For a clearer view of the test strategy, Figure 6.5 shows a block diagram of the setup more from the perspective of logical connectivity. In the FPGA, the GTX resources are used to perform the loop-back BERT test of the external optical fiber links between GTXs from all the FPGA quads. The resources can be invoked by instantiating IP core called IBERT 7 Series GTX from the Vivado IP catalog, which allows users to define and gener-

ate a customized IBERT core to validate the transceivers of the FPGA and other external transceivers. The users can customize the test protocol (a combination of line rate, data width, reference clock, quad count and PLL type), and activated quads by assigning a protocol to each of them. The times that a protocol can be assigned to quads should be exactly the number of the quad count set in the protocol definition. After FPGA implementation, in the case of the protocol defined with just one quad, the IBERT core will try to find all the hardwired data links from the transmitter of one GTX to the receiver of the same GTX just in the current quad that is assigned with the protocol, therefore the maximum number of links that can be found should be four respectively from the four GTXs in this quad. In the case of the protocol defined with more than one quad, hardwired data links can only be found across the quads that are assigned with the protocol and they should go from the transmitter of one GTX to the receiver of its corresponding GTX in another quad. As an example, if the the protocol is defined with two quads, the maximum number of links that can be found should be eight, including the four from one quad to another quad and other four going the other way around. To be more precisely, the IBERT core checks if the communication is successfully established in each of the eight links by looking at the correctness of the received data patterns (usually PRBS) on the receiver side of the link, then just show those links that are checked right. Finally, the IBERT core will perform BERT test on each of the discovered links under the assigned protocols.

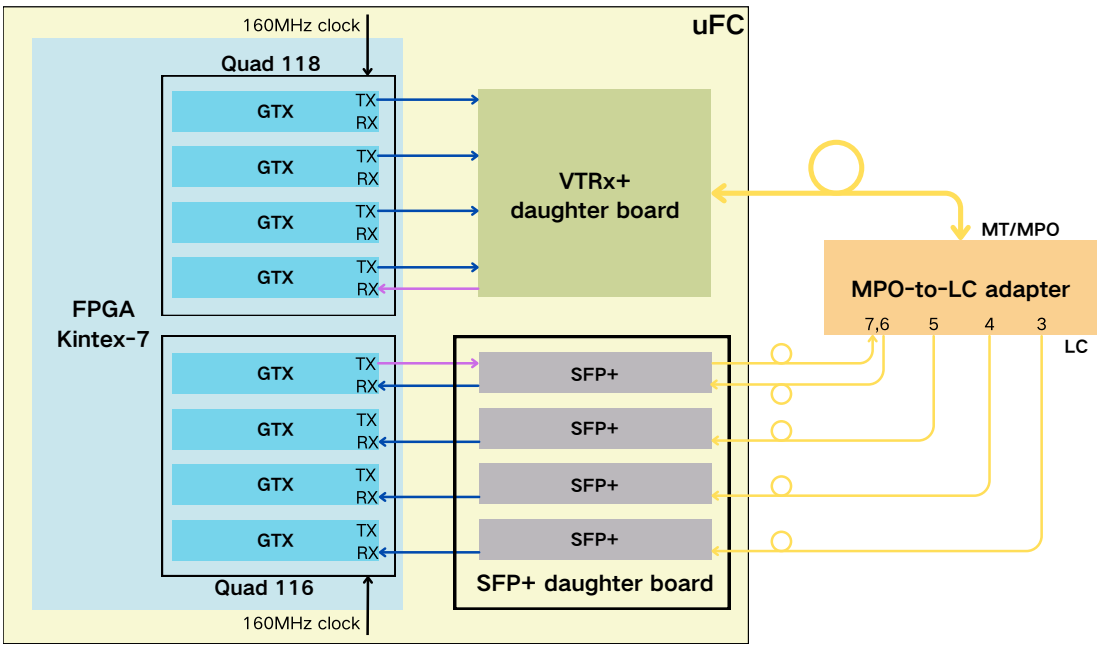
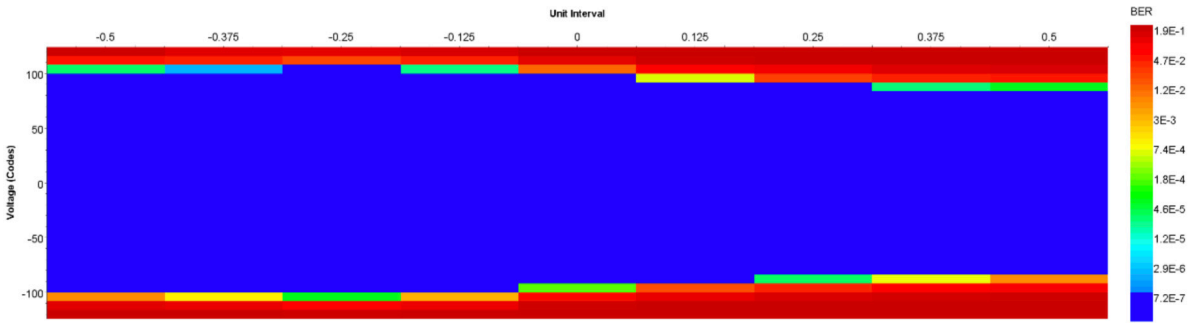


Figure 6.5: Block diagram of the setup from the perspective of logical connectivity.

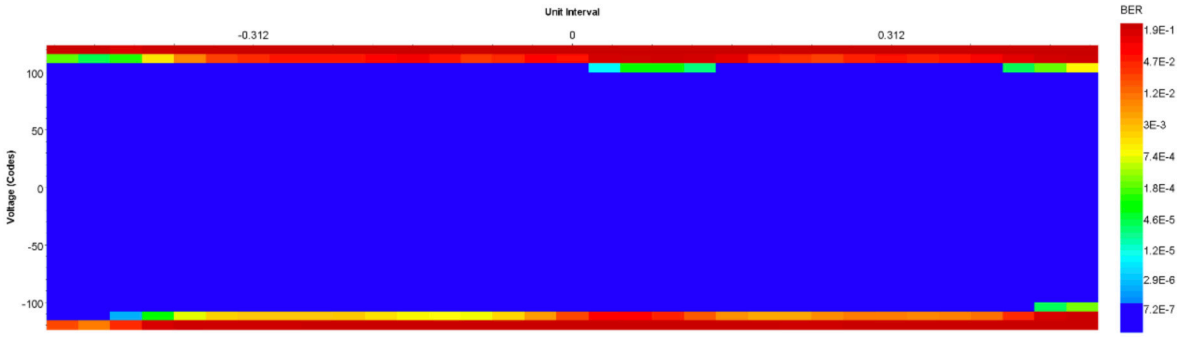
As can be seen from the diagram, the quad 118 and 116 are used to conduct the BERT test, and each of them has four GTXs. The quad 118 is connected to the four uplinks and one downlink of the VTRx+ while the quad 116 is connected to the four downlinks and one uplink

of the SFP+ daughter board with four SFP+ optical transceivers inserted. The reference clocks for the GTXs from both the quads all have the frequency of 160 MHz. In this test, there are two protocols defined with the line rates of 2.56 $Gbit/s$ and 10.24 $Gbit/s$ respectively for the test of the four VTRx+ uplinks and one VTRx+ downlink. The IBERT core will find five hardwired loop-back links (four uplinks and one downlink) then perform 2.56 $Gbit/s$ BERT test on the downlink loop-back link and 10.24 $Gbit/s$ BERT test on the uplink loop-back links. It's worth mentioning that, on the LC connector side of the MPO-to-LC adapter, there are LC connectors labelled with numbers which has one-to-one corresponding relations with the optical fiber inputs on the MT/MPO side, therefore, the fibers on the LC connector side have to be inserted into the correct LC connector slots.

After performing the BERT test on VTRx+ uplinks and downlink, the corresponding bit error rates are calculated and also eye diagrams are made to visualize the transmission performance of those lines. For the 10.24 $Gbit/s$ uplinks, their BERs are all less than 10^{-13}, which can be even better if the BERT test would be conducted for a longer time. Figure 6.6(a) shows an eye diagram of one of the uplinks, which shows a good eye opening. For the 2.56 $Gbit/s$ downlink, its BER is less than 10^{-12}, which may also be better if the BERT test lasts longer. Figure 6.6(b) shows an eye diagram of the downlink, which also shows a good eye opening. It should be pointed out that, during the BERT test, the pre-emphasis and equalization of the GTXs are optimized by using the dedicated Vivado GUI.



(a) Eye diagram of the 10.24 $Gbit/s$ uplink



(b) Eye diagram of the 2.56 $Gbit/s$ downlink

Figure 6.6: Eye diagrams of the VTRx+ uplink and downlink.

6.2.1.3 Test of bPOL12V daughter board

To make sure the power supply of the modular PEB is properly handled, the basic functions of the bPOL12V daughter boards have to be tested, such as the regulated voltage, the efficiency and also the response to temperature variation, especially the high temperature since the DCDC ASIC usually dissipates more heat that greatly increases its temperature.

In order to facilitate the test of the bPOL12V daughter boards with a quite sizable number, a dedicated carrier board is developed for this purpose and it is able to test ten boards at the same time. The design of the carrier board is straightforward. On the board, there are ten pin header connectors that exactly match the connectors on the daughter board. Figure 6.7 shows a picture of the carrier board, where four connectors in the first row are occupied by the bPOL12V daughter boards with shielding cases mounted and other six connectors (one in the first row and five in the second row) are empty. Around each connector slot, there is a dip switch used for setting the 2 control bits and enable signal of each inserted daughter board. There are also ten chassis-mounted wirewound resistors mounted right under the ten connector slots on the other side of the carrier board, and they serve as the variable loads of the bPOL12V daughter boards. For the output voltages, power good signals and temperature signals of the ten daughter boards, they are all routed to the monitoring connectors on the upper-left corner of the carrier board. In addition, on the bottom-left corner, a pair of banana connectors are used for the power supply of the carrier board.

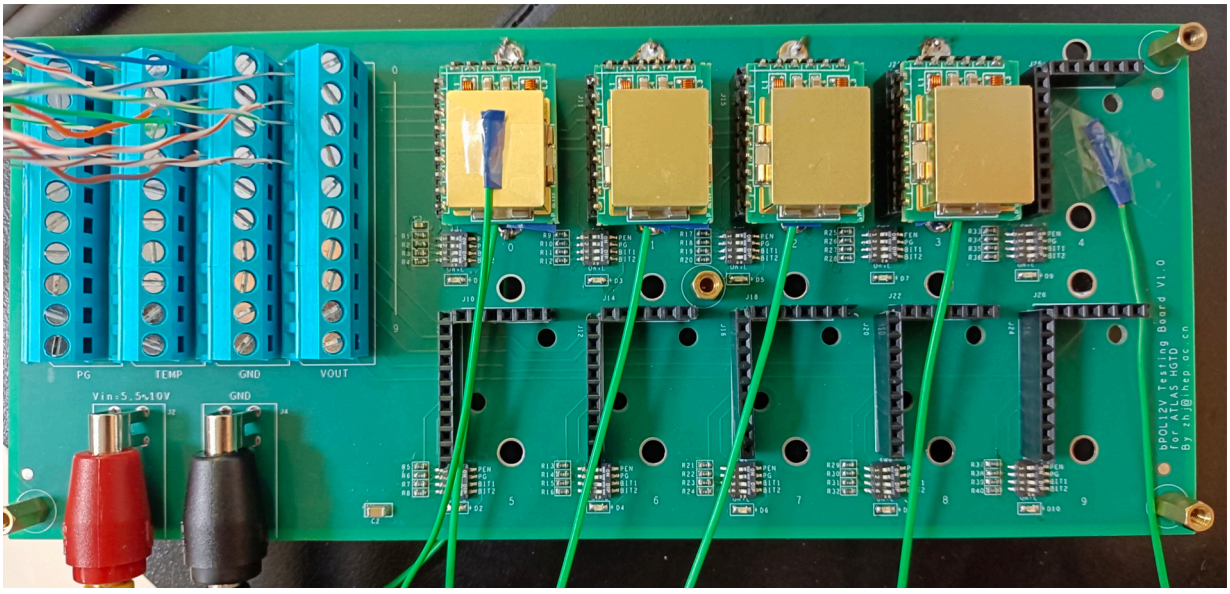


Figure 6.7: Picture of the carrier board for bPOL12V daughter board batch test.

In terms of the efficiency test, the idea is to measure the input and output voltages, also the input and output currents of the bPOL12V daughter board working under the load of a certain resistance. The efficiency is calculated as the ratio of the output power to the input power. Then repeat the procedure by switching the load to another resistance until

the interested range is all scanned. It's found that the efficiency is dependent on the output current. It reaches to the maximum value (about 71%) when the output current is around 2 A. It increases with the output current before 2 A while it decreases with the output current after 2 A. When the output current increases up to 4 A, the efficiency is still above 62%. When the output current is set to be 1 A, the efficiency is about 69%. The result is very similar to the result of ITk DCDC converter, which also uses the bPOL12V power ASIC.

In terms of the thermal test, the idea is to put the carrier board into climate chamber whose temperature varies with time and monitor the response of the bPOL12V daughter boards that are inserted into the carrier board. There are many parameters monitored and sent out of the climate chamber to a multi-channel monitor (GRAPHTEC GL840). The parameters include the output voltages, power good signals and temperature signals from the bPOL12V daughter boards, and also the temperature signals from some measuring temperature probes (thermocouples) mounted on different locations of the setup, such as the surface of the carrier board, the thermal pad and shielding case of the bPOL12V daughter boards, and the wirewound resistors. Figure 6.8(a) shows a picture of the climate chamber where the monitoring wires are routed out and attached to the multi-channel monitor placed on the top of the climate chamber. As can be seen, the power supply unit is also placed on on the top for supplying the carrier board inside the chamber. Figure 6.8(b) shows a picture of the inside view of the climate chamber where the carrier board is placed sideways. The green wires are used to read the temperature probes, the two gray custom-made wires are connected to the monitoring connectors of the carrier board, and the black and yellow wires are used for the voltage input of the carrier board.

(a) Global view of the setup (b) Inside view of the climate chamber

Figure 6.8: Pictures of the thermal test of the bPOL12V daughter boards.

Concerning the test results, firstly it's found that the output voltage of the bPOL12V ASIC decreases as the temperature goes up, but the voltage variation is at the percent level. Figure 6.9 shows a plot about the output voltage as a function of the bPOL12V

thermal pad temperature. As can be seen, the four curves represent the results of four daughter boards, and the values of the output voltages are all very close to the desired voltage ($1.2\ V$). Figure 6.10 shows the relation between the bPOL12V temperature sensor output, also called PTAT (Proportional To Absolute Temperature) voltage, and the its thermal pad temperature. The red curve shows the result from the bPOL12V design group while the other four curves represent the results from four daughter boards. As can be seen, the bPOL12V temperature sensor has a very good linearity. The slopes of the four curves are consistent with the red curve. The variations between the red curve and other four curves are caused by different ways of ASIC temperature measurement while the variations between the four curves are due to the different attachments/fits of the temperature probes and the thermal pads. The reason why the temperature in the Figure 6.10 is less than $100^{\circ}C$ is to avoid possible damage to the ASICs because of high temperature. However, one bPOL12V daughter board is chosen to continue the thermal test by increasing the temperature beyond $100^{\circ}C$. Figure 6.11 also shows the curve of PTAT voltage as a function of the bPOL12V thermal pad temperature but for the chosen daughter board. When temperature reaches up to $120^{\circ}C$, the over-temperature protection of bPOL12V is triggered and the ASIC is immediately shutdown. When the temperature is cool enough (at around $105^{\circ}C$), the ASIC will automatically start to work. If the cooling condition of the system is not improved, the ASIC will probably get stuck in the shutdown-restart cycle, which may cause irreversible damage to the ASIC. Therefore, the lesson learned from this is that a proper cooling for the power supply modules is extremely important.

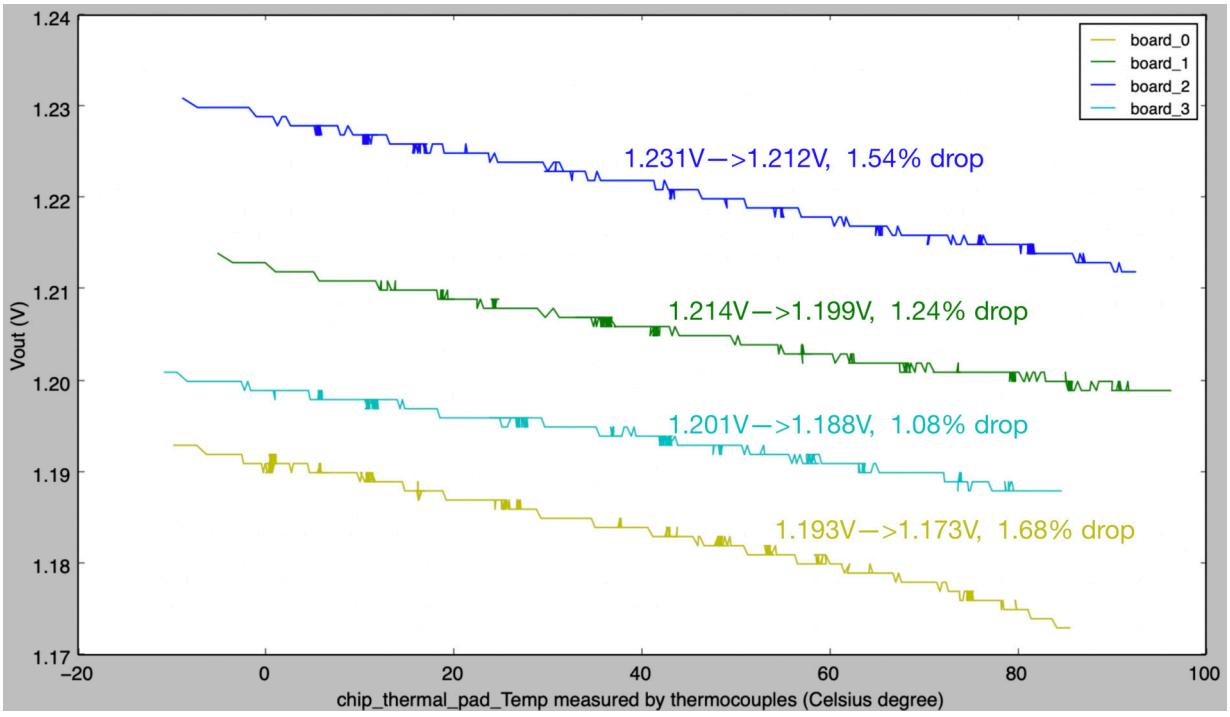


Figure 6.9: Curves of the output voltage as a function of the bPOL12V thermal pad temperature.

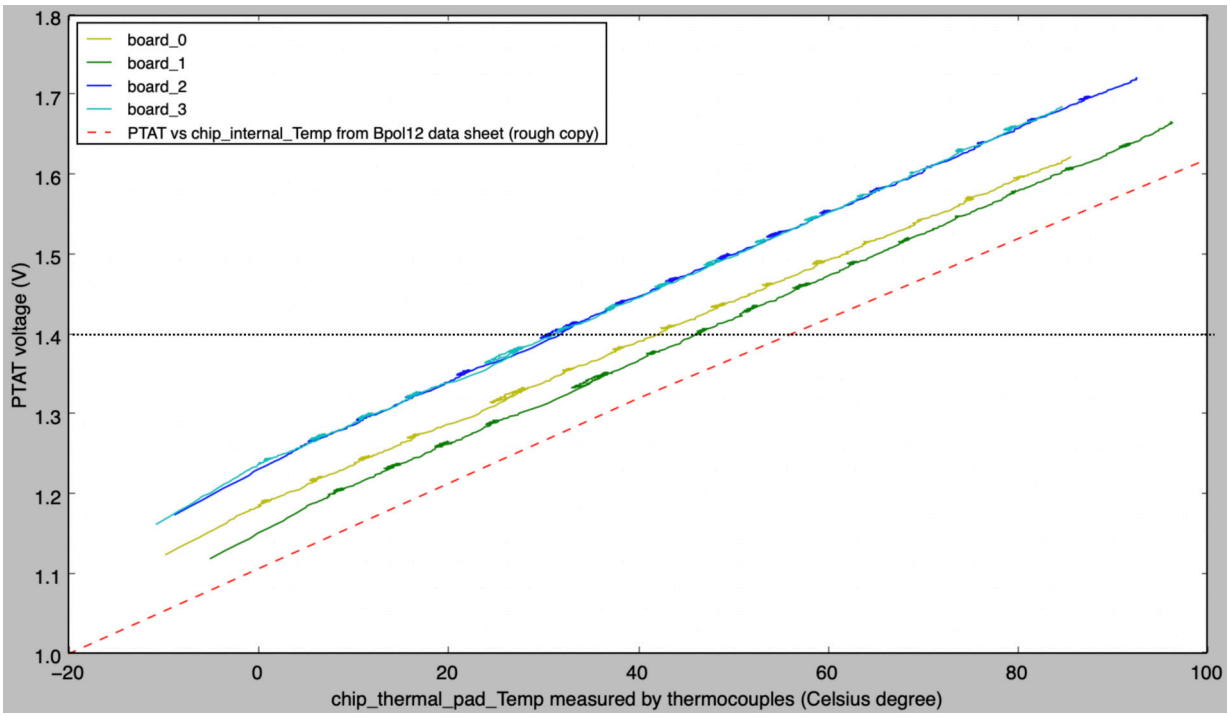


Figure 6.10: Curves of the bPOL12V temperature sensor output as a function of the bPOL12V thermal pad temperature.

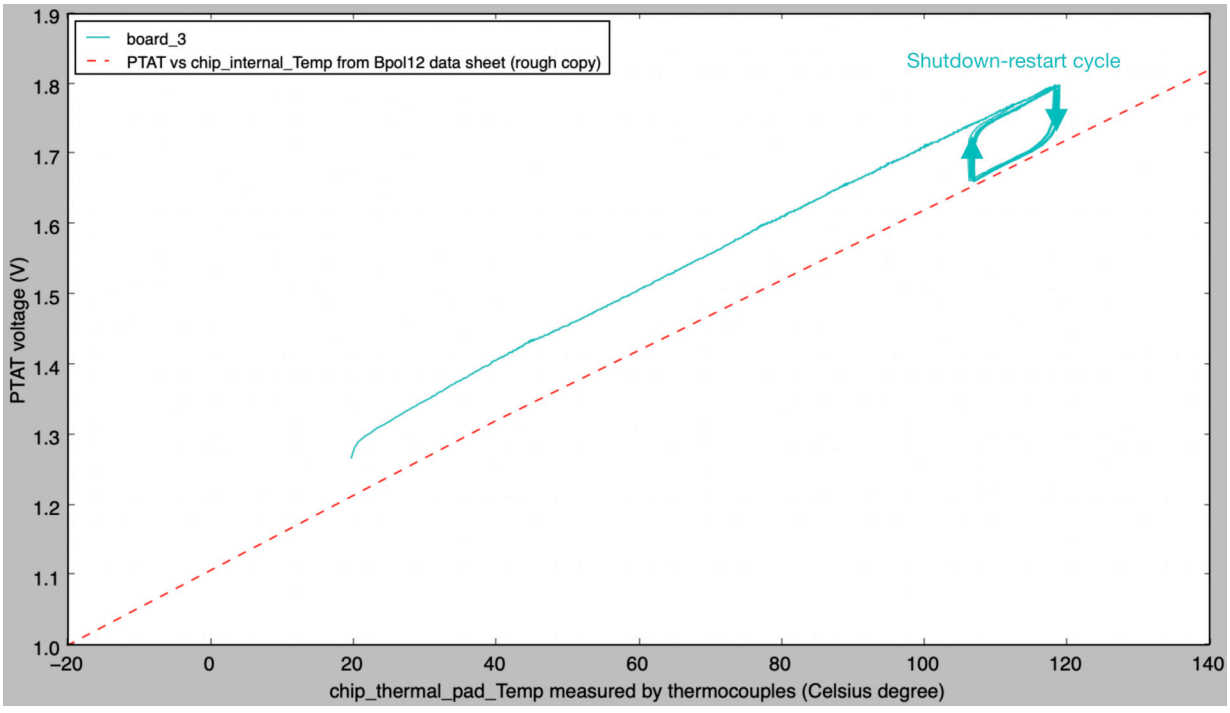


Figure 6.11: The curve of PTAT voltage as a function of the bPOL12V thermal pad temperature with the temperature beyond 100$^{\circ}C$.

In addition, there are also other types of tests that the bPOL12V daughter board has to go through, targeting for the usage on the PEB board. Those tests include reliability test, irradiation test, high-intensity magnetic field test and etc. However, for serving as power supply modules of the modular PEB board, the above-mentioned tests are already adequate. We can say that those bPOL12V daughter boards work as expected from the measurement results.

6.2.2 Test of firmware components

6.2.2.1 Test of the firmware for lpGBT emulator and its DAQ counterpart

The firmware for both lpGBT emulator and its DAQ counterpart has to be tested by implementing it into FPGA before it can be used in the joint test system. Its complexity is quite large so that the resource utilization ratio and power consumption should be evaluated when it's implemented into the FPGA. Even through the feasibility of the firmware has been validated through Vivado simulation, that can not guarantee the real performance of it. Therefore, a special setup is built for this purpose. Figure 6.12 shows a picture of this setup.

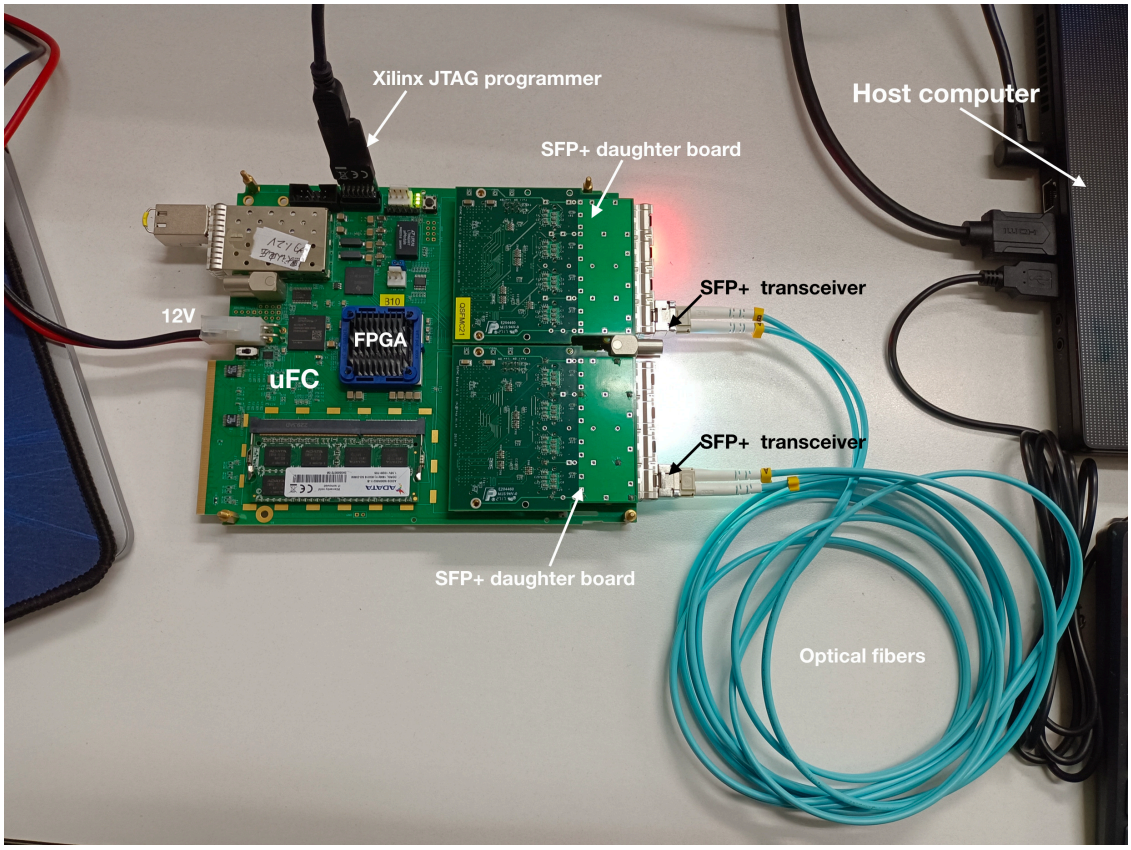


Figure 6.12: Picture of the setup for testing the firmware for lpGBT emulator and its DAQ counterpart.

As shown in the picture, the uFC board is used as the carrier board of this setup.

The FPGA of this board is expected to hold the firmware under the test while the JTAG connector on its upper side is occupied by a Xilinx JTAG programmer that is connected to the host computer through a USB cable. There are two SFP+ daughter boards inserted into the two FMC connectors of the uFC board. Each SFP+ daughter board has four SFP+ connectors, but only of the four connectors is occupied by a SFP+ optical transceiver module that has one optical fiber for data stream input and another one optical fiber for data stream output. Therefore, two optical fibers are used to connect the two transceiver modules, which can be used to establish two data path loops for the lpGBT uplink and downlink. In addition, the uFC board is supplied by a $12\ V$ input voltage through the power input connector on the left side of the board.

Figure 6.13 shows the block diagram of the firmware used for implementing into the uFC FPGA, and the electrical and optical high-speed links. On the left side of the block diagram shows the architecture of FPGA implementation, which is mainly composed of the two pieces of firmwares that require tests. Apart from them, there are also two ILAs used for monitoring the recovered data at the end of the data path. Of course, there are also two data paths (uplink and downlink), both of which start from the FPGA and finally loop back to the FPGA. For the uplink data path, at the beginning, the fixed pattern is fed into the uplink part of the firmware for lpGBT emulator and sent out of the FPGA in the form of $10.24\ Gbit/s$ electrical data stream after the processing of it. Then the electrical data stream goes to the SFP+ optical transceiver and gets converted to the $10.24\ Gbit/s$ optical data stream, which further goes to another SFP+ optical transceiver and gets converted back to the $10.24\ Gbit/s$ electrical data stream. Then it loops back to the FPGA again and gets processed by the uplink part of the firmware for lpGBT DAQ counterpart. Finally, the original fixed pattern is recovered and monitored by the ILA logic, which can display the value of the fixed pattern on the host computer. For the downlink path, data flows in the same way with the uplink data path, so it's not described in details. It's worth noting that the high-speed links between the FPGA and the SFP+ transceiver are electrical while the high-speed links between the two SFP+ transceivers are optical. As can be seen, the data that flows in both the two loop back data paths goes through a quite long journey, including encoding, decoding, serialization, deserialization, elctro-optical conversion, optical-elctro conversion, data width conversion and etc.

After the implementation of the firmware on the Kintex-7 FPGA (XC7K325T) of the uFC, the utilization ratio of the FPGA resources is estimated by the dedicated Xilinx software Vivado. It's found that less than 10% of resources are used. It should be noted that both the two types of firmware are already implemented in the FPGA, and there is still a quite low utilization ratio, which is a very good news for the firmware implementation on the DAQ board (still the uFC board) of the joint test system where just the firmware for lpGBT DAQ counterpart needs to be implemented. Also the total power is also estimated,

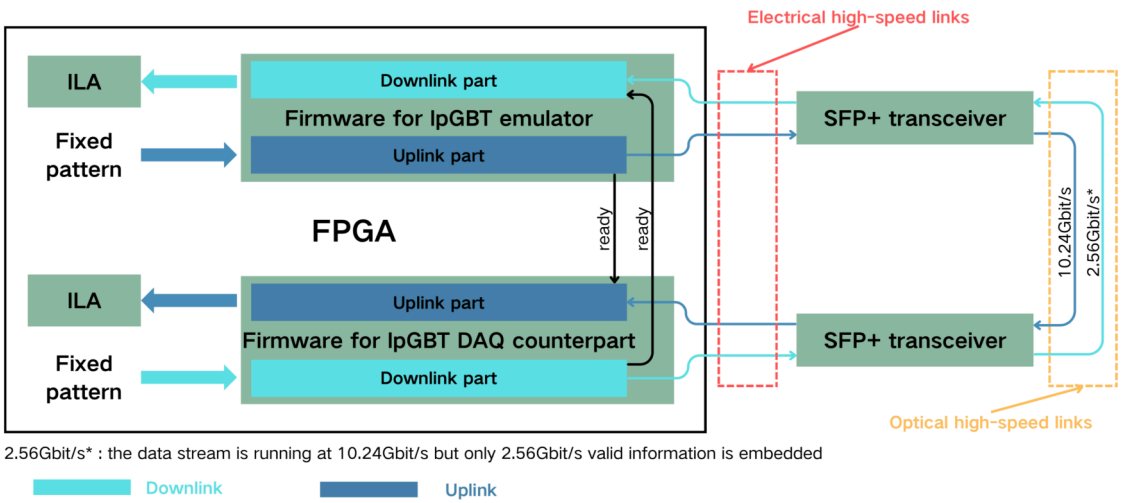


Figure 6.13: Block diagram of the firmware that is used to be implemented into the uFC FPGA, and the electrical and optical high-speed links.

and it's about 1.3 W, which is definitely acceptable. A cooling fan is planned to be used for the uFC board during the joint test, which will accelerate the heat dissipation from the FPGA. Most importantly, the correctness and reliability of the firmwares are verified after performing the loop-back test of the data paths for so many times and always getting the correct recovered fixed patterns at the end of the data path.

6.3 Joint test system setup

So far, all the individual components of the modular PEB are tested and everything is ready. It's time to perform the joint test of the modular PEB system by putting the pieces together into a monolithic whole. The primary goal of the joint test is to verify the hardware design of the modular PEB, so that we are reassured that it can be used for more complicated tests and advanced tasks in the future. To guarantee the correctness and rationality of the hardware design, the joint test has to be performed from the following aspects including system configuration, data transmission, power supply distribution, monitoring, clock and fast command distribution. Among them, the test of data transmission is more complicated and needs to consider versatile link (lpGBT + VTRx+) communication, full data path communication, E-port data rate setting, etc. At the end, we also expect to have answers to our concerns about PEB board design raised up in the Section 5.1.

The joint test system is shown in Figure 6.14. On the left side, there is the fully-equipped carrier board with three FH26W daughter boards inserted on its left three FMC slots, one lpGBT daughter board inserted on the upper-right FMC slot for processing timing data, one lpGBT daughter board on the lower-right FMC slot for processing luminosity data, and one VTRx+ daughter board inserted on the middle-right FMC slot. There are fourteen module emulators connected to the three FH26W daughter boards, among which the upper two

hold five module emulators each while the lower one holds four module emulators due to the limited E-ports of lpGBT. each of the lpGBT daughter board has one UPL board connected for its lpGBT configuration. The pigtail of the VTRx+ is attached to the MPO-to-LC optical fiber adapter through its MPO connector while on the other side of the adapter three optical fibers corresponding to the two uplinks of the two lpGBT and one downlink of the timing (also called DAQ) lpGBT, are routed out and fed into two SFP+ transceivers that are inserted into two slots of the SFP+ daughter board, which are inserted into the upper FMC slot of the uFC board. As can be seen, like in other individual tests, the uFC board is also used as the DAQ board for the modular PEB joint test system. A Xilinx JTAG programmer is connected to the uFC through the JTAG connector located on its lower edge and still it is used for downloading the firmware to the uFC from the host computer. On the upper-left corner of the setup, the power supply unit is used to supply the modular PEB board with a 11 V voltage and the uFC board with a 12 V voltage. In addition, a cooling fan is used to accelerate the heat dissipation of the FPGA on the uFC.

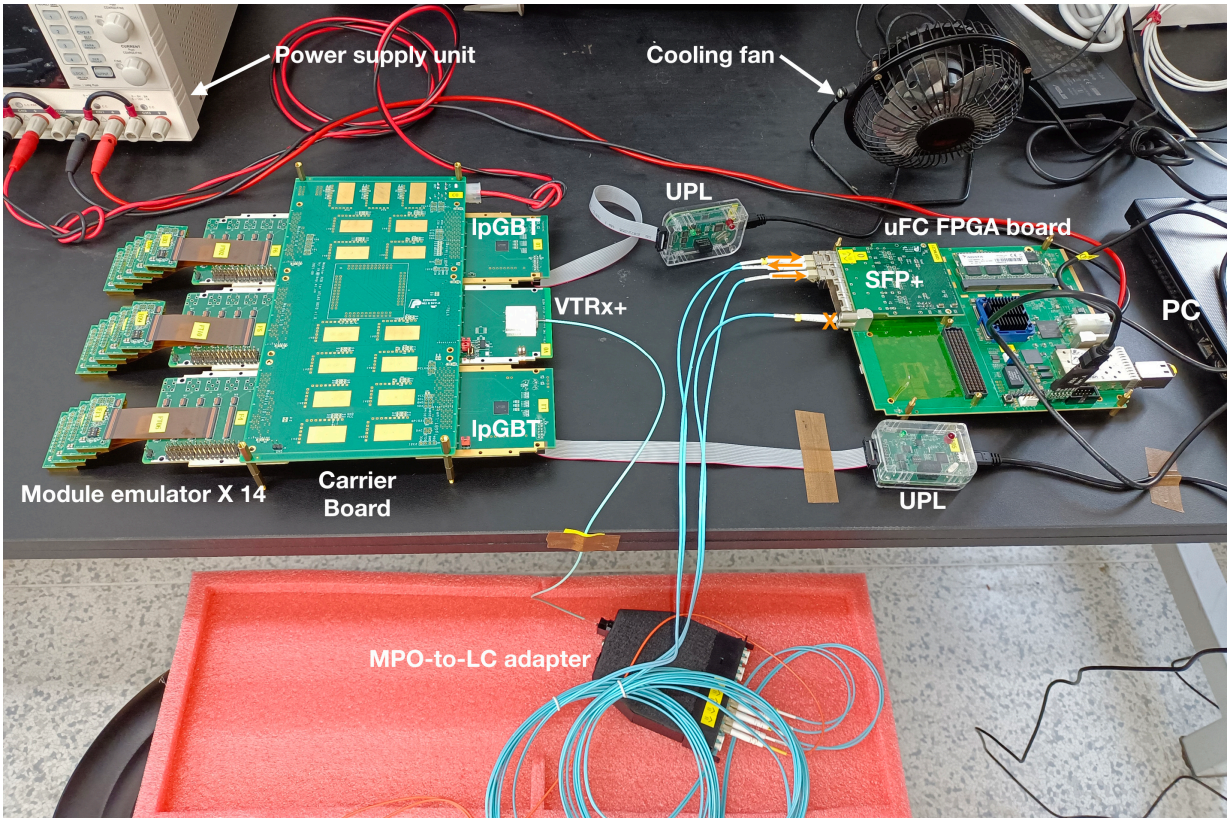


Figure 6.14: Picture of the joint test setup.

It's worth mentioning that the Figure 6.14 shows the final complete setup of the joint test, but there are many simpler versions of setup before the final version, which helps us to have more and more confidence about successfully operating the final version of the modular PEB. For example, the simpler versions may be with just a few module emulators attached or SFP+ daughter board used instead of the VTRx+ daughter board for gradually increas-

ing the complexity of the system and easily debugging possible problems before the system becomes complicated. Therefore, it should be beared in mind that the successful operation of the complete setup is not achieved overnight but from the accumulated experience and lessons learnt from those intermediate versions of setup. In this section, the tests are performed mainly based on the final version of the joint test system, but tests based on the intermediate versions of setup may also be presented.

For the power supply of the module emulators, they are completely from the DCDC converters on the carrier board and delivered by the commercial flat cables once the module emulators are integrated into the system. However, there is also a case when the module emulator should be supplied by other source through the dedicated power supply slots on its board. It happens when the module emulator FLASH memory is programmed, which has to be done before the integration otherwise it will be extremely hard to carry out when they are already attached to the FH26W daughter boards. Once the FLASH memory is programmed, the module emulator will automatically load the firmware from it right after power on. The FLASH memory programming of the module emulator is described in detail in 5.3.3.1.6.

In terms of the lpGBT configuration of the system, two UPLs are used to respectively configure the one timing lpGBT and one luminosity lpGBT. The lpGBT configuration is already described at length in Section 6.2.1.1. The only difference between different lpGBT configurations is just the configuration files used by them. The configuration file stores the desired values of the lpGBT registers that need to be configured, then it gets loaded by the UPL GUI or corresponding script, which will further transfer the file to lpGBT with the help of the UPL board. Finally the lpGBT register values will be updated accordingly thus the working status of the lpGBT changes. Therefore, configuration files should be written based the requirement of the specific testing task. For example, when the data rate setting of the E-ports is required to be changed, in the configuration file, the values of the registers related to that should be modified correspondingly.

In terms of the firmware development for the module emulator and uFC, it varies based on the requirement of the specific testing task, but the main logics should be the same. For example, about the module emulator firmware, two copies of the firmware for module emulator are usually used to respectively serve as the timing data and luminosity data sources. Even though the firmware itself can just generate the timing data, the luminosity lpGBT can of course handle that because technically speaking both the luminosity and timing lpGBTs are the same kind of ASIC but used in different places of the modular PEB. In terms of the firmware development for the uFC, two copies of the firmware for lpGBT DAQ counterpart are usually used to respectively handle the data streams of the two lpGBTs. Since the luminosity lpGBT has just one uplink, for its dedicated firmware copy, only the uplink logic is activated.

For some tests, the FELIX system is also set up to serve as the DAQ system instead of using the uFC board. The FELIX system is the dedicated read-out system developed for the ATLAS experiment and it will be used in the DAQ system of the HGTD. The FELIX setup at our side successively goes through two versions so far. The first version is based on a commercial VC709 board as the carrier board and a server that can support FHFL(Full Height / Full Length) PCIe card. The specified operating system of the server is CentOS 7.9 while the corresponding firmware, software and driver are provided by the ATLAS FELIX Project. The second version is based on the dedicated FELIX card FLX-712-24Ch Nr131 and the same server with the first version. Still the operating system CentOS 7.9 is used but the firmware, software and driver are updated to the latest releases. In addition, different optical fiber adapters are also required based on the version of the FELIX setup. Aided by the FELIX system and its dedicated software, the DAQ of the modular PEB system can be carried out even more easier and efficient.

6.4 Test of system configuration

Generally speaking, the modular PEB configuration is composed of two consecutive steps. The first step is to configure the timing lpGBT while the second step is to configure other components of the system such as the luminosity lpGBT, VTRx+, MUX64, bPOL12V and front-end module emulators. Further, the lpGBT cascade is specifically referring to the control and configuration of the luminosity lpGBT by the timing lpGBT. The primary goal of the modular PEB configuration test is to check the connectivity of the configuration paths from the hardware perspective. Therefore, the test strategy is straightforward, which is to check if the configuration signals can reach their destination along those configuration paths. To carry out this test, the fully-equipped modular PEB joint test system is used. In terms of firmware for this test setup, since there is no data path tested, it just reuses the firmware for the full data path test system. In this case, the only mandatory logic is actually the downlink logic of the firmware for lpGBT DAQ counterpart that provides $2.56\ Gbit/s$ data stream to timing lpGBT for its clock recovery. In terms of the lpGBT configuration, the same configuration files with the full data path test system can still be used.

Figure 6.15 shows the block diagram of the configuration paths in the modular PEB system. The timing lpGBT can be configured either through its I^2C slave port or its built-in serial control logic. In this test, it is configured by UPL through the slave port. At the beginning of its configuration when its PUSM is at the state of ARESET, it will issue a reset pulse RSTOUTB to reset the luminosity lpGBT. After its configuration, the luminosity lpGBT receives the $40\ MHz$ reference clock from it and gets ready to be configured through its slave port that can be accessed by either UPL or the timing lpGBT second I^2C master port, besides it can also be configured by its serial control logic through its EC E-port, which

gets configuration data stream from the timing lpGBT at the data rate of 80 *Mbit/s*. In this test, it is also configured by UPL. Apart from the luminosity lpGBT, other components are also connected to the timing lpGBT. The first I^2C master is used to control the 11th to 14th module emulators; The second one is used to control the 6th to 10th module emulators and the luminosity lpGBT as just mentioned; The third one is used to control the first to 5th module emulators and the VTRx+. The five pins GPIO[15:11] of the timing lpGBT are routed to the fourteen module emulators to provide reset signals for them, so there are cases that several module emulators share one GPIO due to limited number of GPIOs. Another five pins GPIO[10:6] are used to enable the twelve DCDC converters dedicated for module emulator power supply, also there are cases that several DCDC converters share one GPIO. The last six pins GPIO[5:0] are used to select the MUX64 input channels.

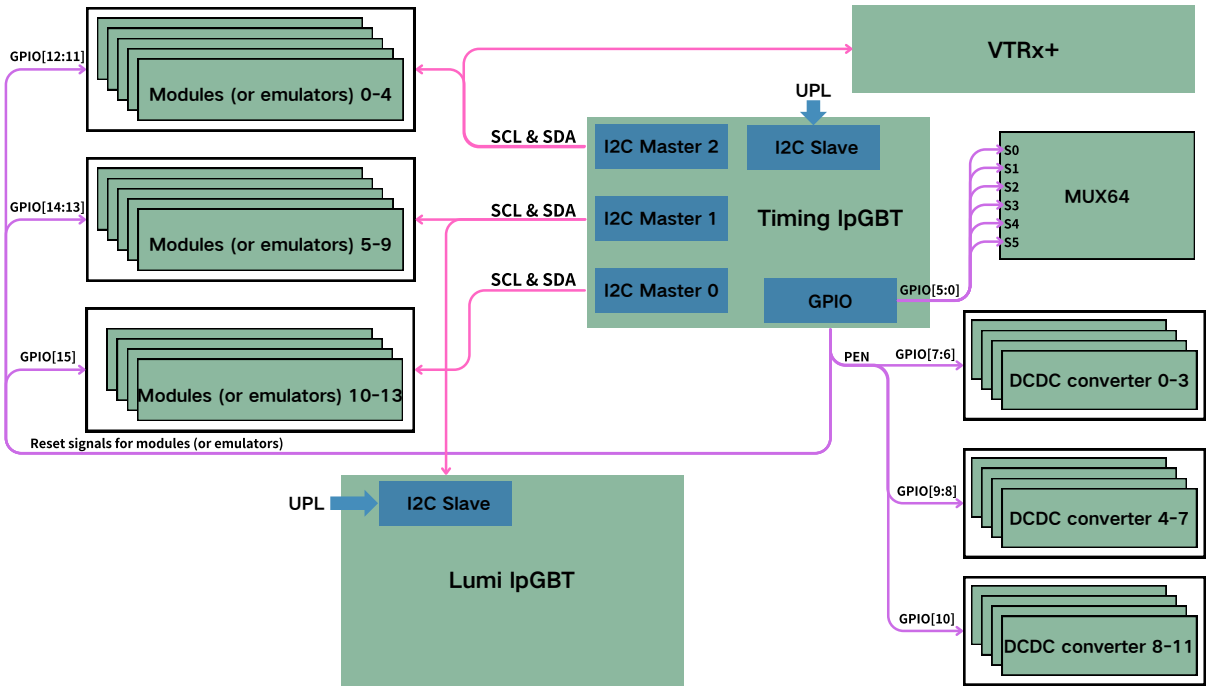


Figure 6.15: Block diagram of the configuration paths in the modular PEB system.

All the above-mentioned configuration paths all go through connectivity test by checking either the configuration of the target components or the arrival of the configuration signal at the end of the path. For example, the configuration paths about DCDC converters are tested by checking if the target DCDC converter is properly enabled. For the I^2C configuration paths about module emulators, the SCL and SDA signals are measured with oscilloscope and checked based the waveforms. The four connections related to the lpGBT cascade including I^2C and EC configuration paths, the 40 *MHz* reference clock and reset pulse RSTOUTB lines, are also tested with oscilloscope. Finally, it's found that all the configuration paths have good connectivity.

6.5 Test of data transmission

6.5.1 Versatile links test

The versatile links refer to the high-speed uplink and downlink between lpGBT and DAQ board. More specifically, the high-speed uplink starts from the lpGBT and flows out at the data rate of $10.24\ Gbit/s$. Then it's fed into the VTRx+ and transferred to optical signal with the same data rate taking the VTRx+ pigtail's optical fiber as its carrier. Finally, the fiber is routed to the DAQ board for further data processing. The high-speed downlink goes the other way around. It starts from the GTX transceiver of the uFC FPGA and directly converted into optical signal by the SFP+ optical transceiver mounted on the uFC board, then flows out at the data rate of $2.56\ Gbit/s$ also taking the optical fiber as its carrier. Then the fiber is routed to the VTRx+ where the optical signal is transferred back into electrical data stream with the same data rate. Finally, the data stream goes to lpGBT where it will be further processed. The versatile uplink and downlink are tested separately.

6.5.1.1 Test of versatile uplink

The test of versatile uplink is to measure the performance of its data transmission. The most commonly-used figure of merit is the eye diagram of the data stream that flows through the uplink. The eye diagram measurement is expected to be performed with the help of the Xilinx IBERT IP core and its corresponding graphical interface embedded in Vivado. There are two versatile uplinks in the system, one is related to the timing lpGBT while the other is related to the luminosity lpGBT. For the experiment setup, the fully-equipped modular PEB joint test system is used.

In terms of the data source of the versatile uplink, there are in total four sources that can be used. One of them is the aggregated data streams coming from the front-end module emulators. Another three of them are from the lpGBT built-in data generators. The three generators are different. The first one is called PRBS7 generator that generates random serialized data streams for all the 28 E-links. The second one is a fixed pattern generator for all the seven E-ports while the third one is also a fixed pattern generator but for the high-speed serializer whose output directly goes to the line driver that can finally drive the serialized data stream out of the lpGBT. Figure 6.16 shows a block diagram of the lpGBT data path pattern generators. On the lower part of the diagram, the pattern generators for the uplink are presented. In this test, each of the last three data sources is selected as the input of the uplink. As a matter of fact, they are actually equivalent to each other regarding the eye diagram measurement since the measurement can always be performed as long as the data source is fed into the uplink.

To configure which data source is used, the related registers like ULDataSource should

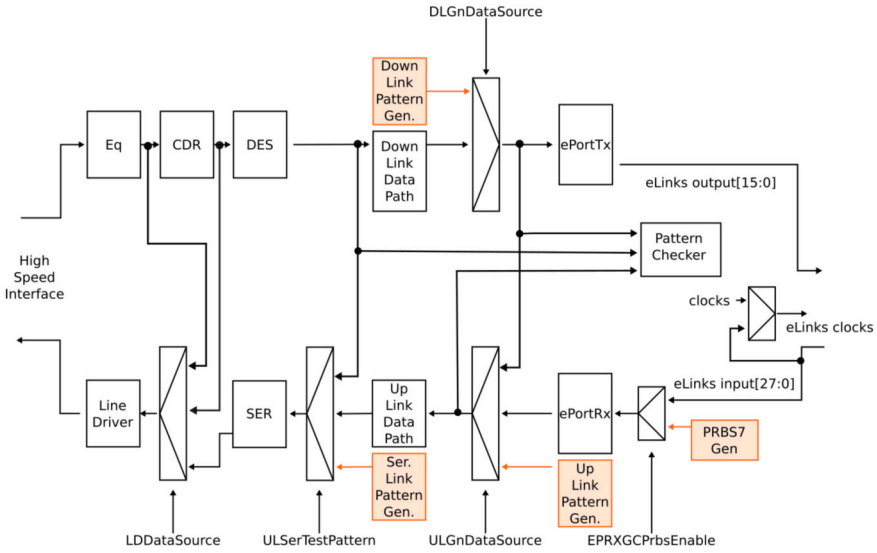


Figure 6.16: Block diagram of the lpGBT data path pattern generators.

be set with certain values. For example, the four LSBs of the register ULDataSource0 control the data input that goes to the high-speed serializer, and it can be either the uplink timing (or luminosity) data or the data patterns from the generator. Actually, there are more than ten kinds of data patterns that the generator can produce, such as PRBS7, PRBS15, constant pattern, clock patterns and etc. The registers related to the line driver pre-emphasis, including LDConfigH and LDConfigL, should be firstly set with proper values, then find optimal ones by scanning the whole parameter ranges. The parameters include the modulation current, duration of the pre-emphasis pulse and pre-emphasis current. Once the data source is set to be timing (or luminosity) data from the frontend, the registers that control the data rates and activated channels of the uplink E-ports should also be set. Those registers include seven control registers for configuring the E-links enabling and data rate of the seven E-ports, 28 registers for configuring the phase selection, level inversion, AC bias, termination and equalization of the 28 uplink E-links. More specifically, for the equalization configuration, four equalization control registers have also to be set. In this test, there is no special requirement on the E-port data rates, therefore, from the perspective of performing test, all the 28 E-links are activated and running at the speed of $320\ Mbit/s$.

For the timing lpGBT, the LOCKMODE pin is set to be 1'b1 to enable reference-less locking and the MODE pins are set to be 4'b1011 to set the lpGBT working as a transceiver with FEC5 and $10.24\ Gbit/s$ high-speed uplink. For the luminosity lpGBT, the LOCKMODE pin is set to be 1'b0 to use the external $40\ MHz$ clock as reference clock and the MODE pins are set to be 4'b1001 to set the lpGBT working as a transmitter also with FEC5 and $10.24\ Gbit/s$ high-speed uplink. Since the second one of the timing lpGBT phase shifter clocks is routed to the luminosity lpGBT as its reference clock, the register PS1Config of the timing lpGBT has to be properly set to configure the frequency, drive strength, fine deskewing and phase shifter delay of the reference clock. It should be noted that the realization of phase

shifter delay configuration also need to set the register PS1Delay, and the output driver of the clock should be set by configure the register PS1OutDriver. For both lpGBTs, the BOOTCNF0 and BOOTCNF1 pins are set to be high to load the configuration information from register array.

In terms of the firmware development for this test, it is quite simple. Figure 6.17 shows its block diagram. The firmware for the module emulator are simply consisted of four OSERDES2 logics that serialize fixed data patterns into four data streams mimicking the two timing data outputs and two luminosity data outputs. On the DAQ side, the two versatile uplinks of the two lpGBTs go to the uFC Kintex-7 FPGA and processed by the IBERT logic instantiated for the quad 116 since the uplinks are routed to the first and second GTXs of this FPGA quad. Finally, the eye diagram measurement of the two uplinks can be performed with the help of the software Vivado. It should be noted that, as described in Section 6.2.1.2, in principle there should be external loop-back wires that connect the receiver and transmitter of each GTX together, but our setup can also work properly as long as the fixed pattern types of the chosen versatile uplink data source is the same with that of the pattern checker used in the IBERT logic since the IBERT pattern checker doesn't care about where the input comes from but do the data pattern check. In this case, we can say that the IBERT pattern checker is somehow deceived by a loop-back link that doesn't come from the corresponding GTX transmitter but from somewhere else. Most importantly, the $2.56\ Gbit/s$ versatile downlink has to be provided to the timing lpGBT high-speed downlink since it is the data source used for recovering primary clock of the lpGBT. To do that, the firmware for lpGBT DAQ counterpart is implemented in the FPGA and only its downlink logic is used to generate the downlink stream from the fixed pattern.

Finally, the transmission performance of the versatile uplinks is evaluated with the setup. Figure 6.18 shows the eye diagram measurement of the versatile uplink related to the timing lpGBT. As can be seen, it already shows a good performance even without performing a thorough parameter scan of the lpGBT high-speed line driver pre-emphasis, which is currently set with some relatively suitable values. The eye diagram of the other link has the similar performance and it's not presented here. In addition, the bit error rates are also measured for the two links and both of them are less than 1×10^{-14}.

The pre-emphasis parameter scan of the VTRx+ uplink is carried out by measuring the eye diagram of the VTRx+ output optical signal through an optical probe and oscilloscope. The idea is to fine-tune corresponding parameters until the best eye diagram opening is achieved. There are two indicating factors that can evaluate the eye diagram opening, the eye height and eye width. The parameters include bias current varying from 0 to 127, modulation enable, modulation current varying from 0 to 127, pre-emphasis amplitude varying from 0 to 7, rising edge pre-emphasis enable and falling edge pre-emphasis enable. The modification of those parameters is conducted by I^2C communication that is issued from

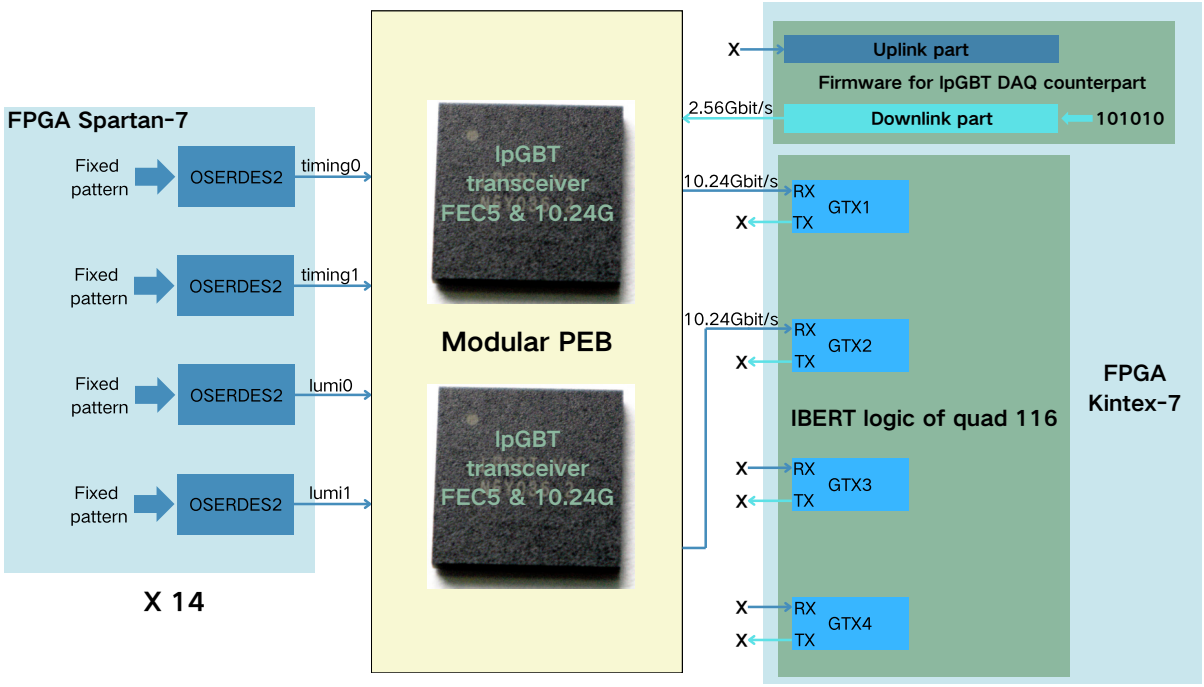


Figure 6.17: Block diagram of the firmware for versatile uplink test.

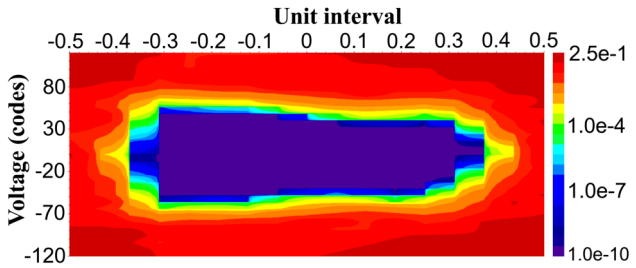


Figure 6.18: Eye diagram of the versatile uplink related to the timing lpGBT.

FELIX side and further broadcasted to the VTRx+ through the I^2C bus. Finally, a set of optimal parameters is selected and the eye height and eye width are respectively larger than 90 μW and 55 ps. The optimal parameters will be adopted for initial configuration in the future experiments. It's worth mentioning that, in this setup, the transmission line between lpGBT and VTRx+ already has a large length of about 21 cm due to the modular design of the modular PEB.

6.5.1.2 Test of versatile downlink

The test of versatile downlink is to measure the performance of its data transmission. Just like the versatile uplink, it also uses the eye diagram as the figure of merit. However, instead of using the Xilinx IBERT IP core to carry out the eye diagram measurement, the lpGBT built-in logic called EOM (eye opening monitor) is used cooperatively with UPL toolkit for that purpose. There is just one versatile downlink in the system and it is related to the timing lpGBT. For the experiment setup, the fully-equipped modular PEB joint test system is used.

The EOM circuit allows the user to perform an on-chip eye diagram measurement of the incoming 2.56 $Gbit/s$ high-speed downlink data. The circuit is used right after the equalizer of the high-speed downlink as depicted in Figure 6.16. The 2.56 $Gbit/s$ data is firstly compared with a static voltage using a differential comparator whose output is sampled with a phase interpolated clock and stored in a 16-bit counter. The value of the counter can be acquired by reading the corresponding lpGBT registers called EOMCouterValueH and EOMCouter-ValueL for storing eight MSBs and eight LSBs of the counter value. Concerning the output of the comparator, if the static voltage is within the high-speed downlink voltage limits, it will toggle (eye opening), which leads to increment of the counter. Otherwise, it will be either static 0 or 1, which will not change the counter value. The x-axis is the time, which can be scanned by adjusting the phase interpolated clock whose sampling phase is controlled by the register EOMConfigL. The y-axis is voltage, which can be scanned by adjusting the static voltage of the comparator whose value is controlled by the register EOMvofSel. For every (x,y) point, the measurement is performed within a certain amount of time whose duration is controlled by the register EOMConfigH, and during the time, the count of the eye opening is recorded by the counter.

In terms of the firmware of this test system, the same firmware developed for the versatile uplink test system as shown in Figure 6.17 is reused. The only difference is that the IBERT logic is not required any more then deactivated in this test. As can be seen, the data source of the versatile downlink is the fixed pattern then fed into the dowlink logic of the firmware for lpGBT DAQ counterpart. In terms of the lpGBT configuration, it is basically using the same configuration files with the versatile uplink test system except that the registers related to the eye opening monitor have to be handled with a specific procedure to be able to

fulfill the eye diagram measurement flow. Generally speaking, the eye diagram is measured by repeatedly setting the EOM parameters (i.e. setting the scan point on the voltage-time plane) and reading the 16-bit counter (i.e. acquiring the eye opening times at the certain point) with the help of the UPL toolkit. To do this, a dedicated button is developed for the GUI to automatically handle the complex measurement flow. After the execution of the operations behind the button, an eye diagram of the 2.56 *Gbit/s* high-speed downlink will be created, and it is shown in Figure 6.19. As can be seen, it already shows a very good eye opening even without performing a thorough parameter scan of the high-speed downlink equalizer yet in this test. The currently-used parameters of the equalizer are just set with relatively suitable values.

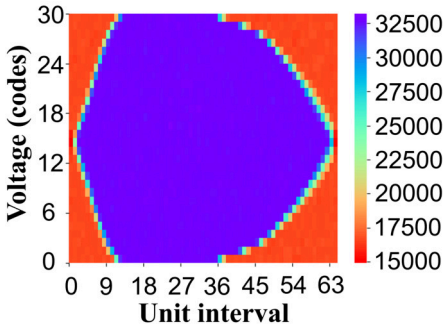


Figure 6.19: Eye diagram of the versatile downlink related to the timing lpGBT.

The full data path test includes the test of both the uplink data path and downlink data path. For the uplink data path, it starts all the way from the module emulators to the FPGA of the DAQ board. For the downlink path, it goes in the opposite way, which is from the FPGA of the DAQ board to the module emulators. The idea is to test the performance of the two paths by checking the correctness of the received data at the end of each path. For the experiment setup, the fully-equipped modular PEB joint test system is used. The firmware for the test system is described in Figure 6.21. There are actually two uplink data paths (timing and luminosity) and one downlink data path in the system.

For the timing data path, the data source is from the ALTIROC logics for timing data that is implemented in the module emulator FPGA, then the data is serialized by the OSERDES2 logic. Each module emulator sends out two serialized timing data streams, which means that 28 timing data streams of the 14 module emulators go to the timing lpGBT and get aggregated into a 10.24 *Gbit/s* high-speed data stream. Then the high-speed data stream is processed by the uplink part of the lpGBT DAQ logic and recovered to the intermediate format before the lpGBT processing. Finally, the intermediate format is totally recovered to the original data format before the ALTIROC processing and goes to the ILA logic with which the received data can be displayed on the host computer. Also the received data is checked by a pattern checker whose output is fed into the ILA as well.

For the luminosity data path, the data is generated in a simple way in the module

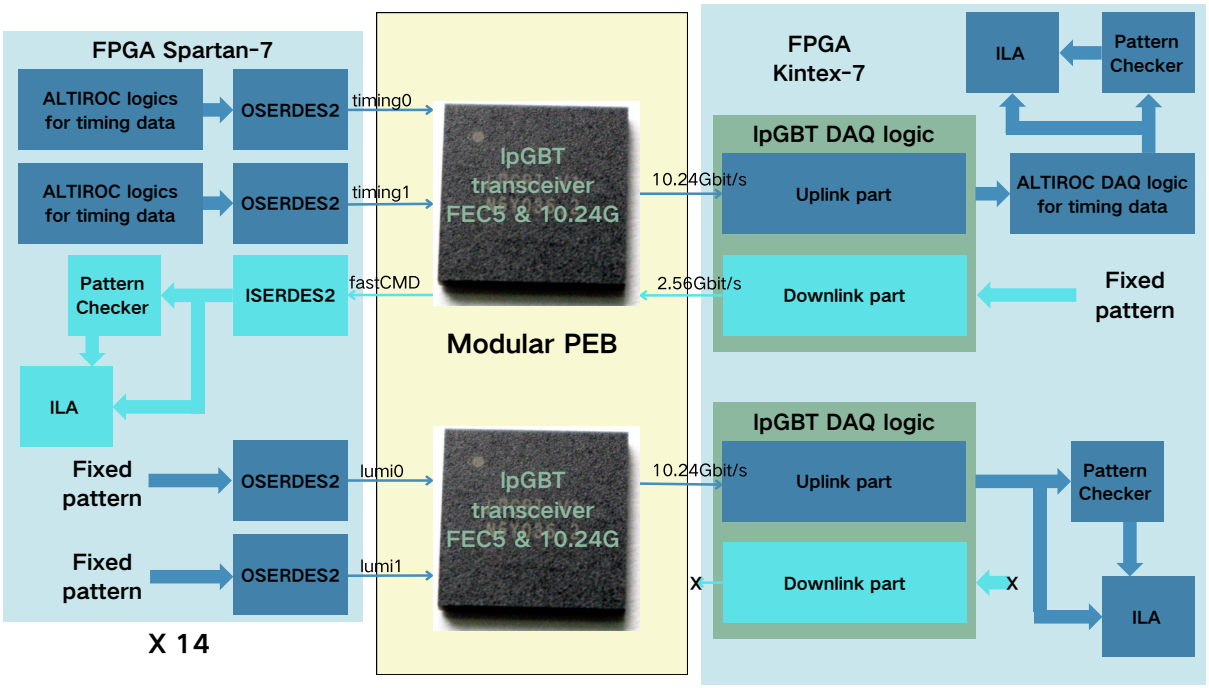


Figure 6.20: Block diagram of the firmware for the test system.

emulator FPGA. As can be seen, the fixed patterns are serialized by the OSERDES2 logic and sent out of the emulator as pseudo luminosity data streams. Just like the timing data stream, there are also two luminosity data streams going out for each module emulator, but they are routed to the luminosity lpGBT. After the processing of the lpGBT, all the 28 serialized luminosity data streams from the 14 module emulators are also aggregated into a 10.24 *Gbit/s* high-speed data stream, which is further handled by the uplink part of another copy of lpGBT DAQ logic. Finally the original fixed patterns are recovered and go to the ILA logic for user to monitor. Also the received fixed patterns are checked by the pattern checker result is sent to the ILA as well. It should be noted that the downlink part of this copy of lpGBT DAQ logic is not activated since it interacts with the luminosity lpGBT that works in the simplex transmitter mode.

For the downlink data path, it goes the other way around. For simplicity, its data source is from the fixed data pattern, which is then sent to the downlink part of the lpGBT DAQ logic and gets out of the FPGA in the form of a 2.56 *Gbit/s* serialized high-speed data stream. After the processing of the timing lpGBT, it is distributed to the sixteen downlink E-links of the four downlink E-ports, then fourteen of them are routed to the fourteen module emulators. Finally, the fourteen downlink data streams are deserialized by the ISERDES2 logics into original fixed pattern, which is monitored by the ILA logic. In addition, the received fixed pattern is checked by the pattern checker whose output also goes to the ILA logic for the user to monitor. It's worth mentioning that the block diagram of the firmware mainly describes the test system from the aspect of FPGA logic, some details about hardware may be omitted. For example, between the lpGBT and the uFC FPGA,

there are VTRx+ and SFP+ optical transceivers; There are also three FH26W daughter boards as the interface boards between module emulators and the modular PEB carrier board. Besides, inside the uplink logic and downlink logic of the firmware for lpGBT DAQ counterpart, the GTX logic is already included for handling the high-speed data streams, therefore, it is not explicitly presented in the block diagram.

In terms of the lpGBT configuration, both of the timing and luminosity lpGBTs are basically configured in the same way with the test system for versatile uplink described in Section 6.5.1.1. The main difference is the configuration on the lpGBT uplink E-ports and downlink E-ports. From the perspective of performing test, all the lpGBT uplink and downlink E-links should be used for data transmission with this test system. Therefore, for the uplink E-links of the two lpGBTs, they are configured to work in the Pattern I shown in Table 5.1, which means that all the E-ports are set to be $320\ Mbit/s$ and all the E-links of each E-port will be working at the data rate of $320\ Mbit/s$. For the downlink E-links of the timing lpGBT, in order to enable all of them, the mirror function described in Section 5.2.2 is activated. The register EPTXControl is used to enable the mirror function of the four E-ports. The first E-link of each E-port is only activated and set to work at the data rate of $320\ Mbit/s$ by configuring the registers EPTX10Enable, EPTX32Enable and EPTXChnCntr where the other parameters like output driver and pre-emphasis can also be adjusted. With the activation of the mirror function for the four E-ports, all the downlink E-links will be working at the speed of $320\ Mbit/s$ and the E-links from the same E-port hold the same data streams.

With the experiment setup, firmware implementation and lpGBT configuration all done, the correctness of all the data paths are checked. It's found that, the communications on the downlink data path and two uplink data paths are successfully established and all work as expected. Thanks to pattern check and ILA logic used at the end of each data path, the bit error rate of each path is also calculated as the ratio of the correctly-received data bits to all the received data bits at the end of the data path. The value is always less than 1×10^{-13}, which is already good for data transmission even without performing a thorough parameter scans of the pre-emphasis and equalizer along the paths. The currently-used values for those parameters are just relatively suitable.

6.5.2 Full data path test

The full data path test includes the test of both the uplink data path and downlink data path. For the uplink data path, it starts all the way from the module emulators to the FPGA of the DAQ board. For the downlink path, it goes in the opposite way, which is from the FPGA of the DAQ board to the module emulators. The idea is to test the performance of the two paths by checking the correctness of the received data at the end of each path. For the experiment setup, the fully-equipped modular PEB joint test system is used. The

firmware for the test system is described in Figure 6.21. There are actually two uplink data paths (timing and luminosity) and one downlink data path in the system.

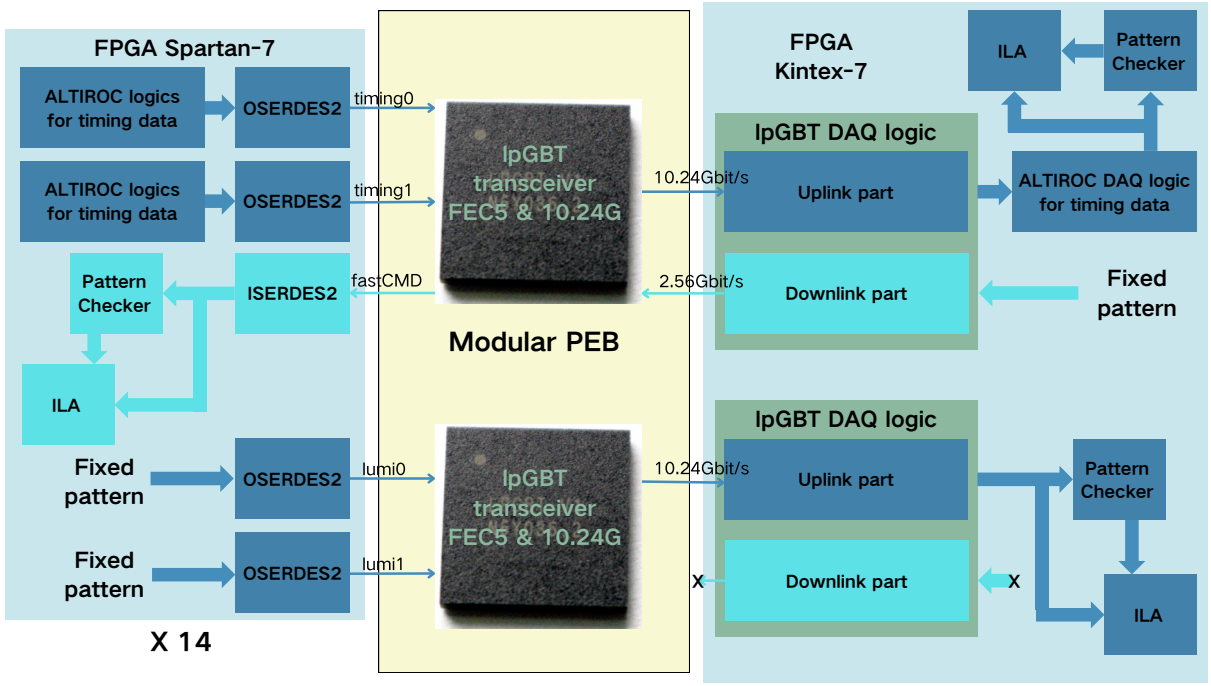


Figure 6.21: Block diagram of the firmware for the test system.

For the timing data path, the data source is from the ALTIROC logics for timing data that is implemented in the module emulator FPGA, then the data is serialized by the OSERDES2 logic. Each module emulator sends out two serialized timing data streams, which means that 28 timing data streams of the 14 module emulators go to the timing lpGBT and get aggregated into a 10.24 *Gbit/s* high-speed data stream. Then the high-speed data stream is processed by the uplink part of the lpGBT DAQ logic and recovered to the intermediate format before the lpGBT processing. Finally, the intermediate format is totally recovered to the original data format before the ALTIROC processing and goes to the ILA logic with which the received data can be displayed on the host computer. Also the received data is checked by a pattern checker whose output is fed into the ILA as well.

For the luminosity data path, the data is generated in a simple way in the module emulator FPGA. As can be seen, the fixed patterns are serialized by the OSERDES2 logic and sent out of the emulator as pseudo luminosity data streams. Just like the timing data stream, there are also two luminosity data streams going out for each module emulator, but they are routed to the luminosity lpGBT. After the processing of the lpGBT, all the 28 serialized luminosity data streams from the 14 module emulators are also aggregated into a 10.24 *Gbit/s* high-speed data stream, which is further handled by the uplink part of another copy of lpGBT DAQ logic. Finally the original fixed patterns are recovered and go to the ILA logic for user to monitor. Also the received fixed patterns are checked by the pattern checker result is sent to the ILA as well. It should be noted that the downlink part of this

copy of lpGBT DAQ logic is not activated since it interacts with the luminosity lpGBT that works in the simplex transmitter mode.

For the downlink data path, it goes the other way around. For simplicity, its data source is from the fixed data pattern, which is then sent to the downlink part of the lpGBT DAQ logic and gets out of the FPGA in the form of a 2.56 *Gbit/s* serialized high-speed data stream. After the processing of the luminosity lpGBT, it is distributed to the sixteen downlink E-links of the four downlink E-ports, then fourteen of them are routed to the fourteen module emulators. Finally, the fourteen downlink data streams are deserialized by the ISERDES2 logics into original fixed pattern, which is monitored by the ILA logic. In addition, the received fixed pattern is checked by the pattern checker whose output also goes to the ILA logic for the user to monitor. It's worth mentioning that the block diagram of the firmware mainly describes the test system from the aspect of FPGA logic, some details about hardware may be omitted. For example, between the lpGBT and the uFC FPGA, there are VTRx+ and SFP+ optical transceivers; There are also three FH26W daughter boards as the interface boards between module emulators and the modular PEB carrier board. Besides, inside the uplink logic and downlink logic of the firmware for lpGBT DAQ counterpart, the GTX logic is already included for handling the high-speed data streams, therefore, it is not explicitly presented in the block diagram.

In terms of the lpGBT configuration, both of the timing and luminosity lpGBTs are basically configured in the same way with the test system for versatile uplink described in Section 6.5.1.1. The main difference is the configuration on the lpGBT uplink E-ports and downlink E-ports. From the perspective of performing test, all the lpGBT uplink and downlink E-links should be used for data transmission with this test system. Therefore, for the uplink E-links of the two lpGBTs, they are configured to work in the Pattern I shown in Table 5.1, which means that all the E-ports are set to be 320 *Mbit/s* and all the E-links of each E-port will be working at the data rate of 320 *Mbit/s*. For the downlink E-links of the timing lpGBT, in order to enable all of them, the mirror function described in Section 5.2.2 is activated. The register EPTXControl is used to enable the mirror function of the four E-ports. The first E-link of each E-port is only activated and set to work at the data rate of 320 *Mbit/s* by configuring the registers EPTX10Enable, EPTX32Enable and EPTXChnCntr where the other parameters like output driver and pre-emphasis can also be adjusted. With the activation of the mirror function for the four E-ports, all the downlink E-links will be working at the speed of 320 *Mbit/s* and the E-links from the same E-port hold the same data streams.

With the experiment setup, firmware implementation and lpGBT configuration all done, the correctness of all the data paths are checked. It's found that, the communications on the downlink data path and two uplink data paths are successfully established and all work as expected. Thanks to pattern check and ILA logic used at the end of each data path, the

bit error rate of each path is also calculated as the ratio of the correctly-received data bits to all the received data bits at the end of the data path. The value is always less than 1×10^{-13}, which is already good for data transmission even without performing a thorough parameter scans of the pre-emphasis and equalizer along the paths. The currently-used values for those parameters are just relatively suitable.

The bit error rate tests of the uplink and downlink E-links are also performed. For the uplink E-link, the fixed data pattern (PRBS7) is firstly generated by the FPGA logic of module emulator and sent out to the lpGBT uplink E-port through the $70\ cm$ FLEX tail and wires on the modular PEB board, finally it gets checked by the lpGBT built-in pattern checker. From the correctness of those received data patterns, the bit error rate of the data transmission in the uplink E-link can be calculated. The maximum data rate $1.28\ Gbit/s$ is tested, and its related bit error rate is less than 5×10^{-13}. For the downlink E-link, the data flows in the opposite direction, which is starting from the lpGBT built-in downlink pattern generator (PRBS7), going through the wires on the modular PEB board and the $70\ cm$ FLEX tail and arriving at the pattern checker implemented in the FPGA logic of the module emulator. The working data rate ($320\ Mbit/s$) of downlink E-link is tested and the bit error rate is less than 1×10^{-12}.

6.5.3 Test of lpGBT uplink E-port data rate setting

The lpGBT uplink E-port can support three kinds of data rates, $320\ Mbit/s$, $640\ Mbit/s$ and $1280\ Mbit/s$. Taking into account the fact that there are up to seven uplink E-ports, many data rate setting patterns can be set for those E-ports as described at length in Section 5.2.3, and the typical data rate patterns are listed in Table 5.1. The idea of the test is to check the feasibility of each data rate pattern. More specifically, the correctness of the received data at the end of the uplink data path will be checked with the uplink E-ports configured with the selected data rate pattern. In terms of the experiment setup, the fully-equipped modular PEB joint test system is used. In terms of the firmware, it is exactly the same with that of the full data path test system.

In terms of the lpGBT configuration, basically, it uses the same configuration files with that of the full data path test system except the registers about uplink E-ports configuration. Every time the data rate patterns like I, II, III, IV and others are successively tested, the registers should be adjusted accordingly. There are seven dedicated EPRXControl registers for configuring the E-links enabling and data rate of the seven E-ports. In addition, there are also 28 EPRXChnCntr registers corresponding to the 28 uplink E-links for configuring their phase selection, level inversion, AC bias, termination and equalization. In order to perform a complete equalization configuration, the four EPRXEqControl registers should also be set. For the pattern I, seven E-ports are set to be $320\ Mbit/s$ and all the E-links of each E-port are enabled, so the two timing or luminosity data streams of each module

emulator work at the data rate of 320 $Mbit/s$. For the pattern II, seven E-ports are set to be 640 $Mbit/s$ and just the first and second E-links of each E-port are enabled, so the two timing or luminosity data streams of seven module emulators work at the data rate of 640 $Mbit/s$ and the other seven module emulators are deactivated. For the pattern III, the first six E-ports are set to be 1.28 $Gbit/s$ and just the first E-link of each E-port is enabled, so the two timing or luminosity data streams of just three module emulators work at the data rate of 1.28 $Gbit/s$; the last E-port is set to be 320 $Mbit/s$ and all the E-links of it are enabled, so the two timing or luminosity data streams of the last two module emulators work at the data rate of 320 $Mbit/s$; other module emulators are deactivated. For the other data rate patterns, they are not depicted here.

The data rate patterns are tested one by one on the timing and luminosity lpGBTs. The results show that the received data at the end of the uplink data path are always correct no matter which data rate pattern is set, which means all the data rate pattern show good performance and can be used in the future test.

6.6 Test of power supply

The output ripple of the bPOL12V output is an important parameter for the ALTIROC ASIC, which requires its value as small as possible. With the latest bPOL12V daughter board design, the output ripple can be less than $10mV$ with the output voltage of $1.2V$, the output current of $3A$ and the working temperature of $20^{\circ}C$.In terms of the efficiency test, it's found that the efficiency is dependent on the output current. It reaches to the maximum value when the output current is around 1.5 A. It increases with the output current before 1.5 A while it decreases with the output current after 1.5 A. The efficiency can reach to around 55% with the output voltage of $1.2V$, the output current of $3A$ and the working temperature of $20^{\circ}C$.

The power consumptions for both lpGBT and VTRx+ are measured based on the full-equipped joint test system with fourteen module emulators attached. The setup is placed into climate chamber to measure the power consumption variations with temperature. The working temperature of the climate chamber gradually goes up from $-40^{\circ}C$ to $60^{\circ}C$. It's found that, across the whole working temperature range, the power consumptions of both lpGBT and VTRx+ increase slightly (by percent level) as temperature goes up. The power consumption of the timing lpGBT is larger than that of the luminosity lpGBT, the former has the maximum value of around 0.4 W while the latter has the maximum value of around 0.3 W. For the VTRx+, the total power consumption can reach up to 0.18 W.

6.7 Test of monitoring network

The monitoring network of the modular PEB system has been already depicted at length in Section 5.2.5. Figure 5.5 shows an overview of the monitoring strategy. The main purpose of the monitoring network is to monitor important voltages, currents, temperatures and other key signals of the modular PEB components including the five signals of each module emulator, power status and temperatures of DCDC converters, temperature and RSSI of the VTRx+, power supply and temperature of the carrier board. For this test, the experiment setup and firmware are also the same with the full data path test. The lpGBT configuration is also the same with that of the full data path test except the registers related to ADCs and GPIOs, which will be used during the test.

The monitoring network test is to check the connectivity and functionality of each monitoring path in this network. The idea is to read the signals to be monitored at the end of the monitoring paths, which are the ADCs and GPIOs of the timing and luminosity lpGBTs. More specifically, to read the signal that reaches a GPIO pin, the pin has to be set as an input and its state stored in a dedicated register can be read by the UPL; to read the signal that is connected to an ADC pin, this pin has to be selected as the input of the MUX of ADC logic and the converted values are stored in dedicated registers whose value can also be accessed by the UPL. Based on the correctness of the values read by the UPL, we can deduce the connectivity and functionality of each monitoring path. For the monitoring paths that monitor the temperatures of the front-end modules, the module emulators use static voltages to mimic the voltage outputs from ALTIROC temperature sensors.

Concerning the registers related to the ADCs, the register ADCConfig is used to activate the ADC logic and set the gain for the ADC differential amplifier while the register ADCSelect is used to select the input channels of the ADC's positive and negative input MUXes. The input channels can be selected from not only the eight ADC pins but also internal sources like the DAC output, some core voltages and temperature sensor output. The two read-only registers ADCStatusH and ADCStatusL are used to record the status of the ADC and the result of last conversion. Concerning the registers related to the GPIOs, there are the PIODirH and PIODirL for controlling the direction of the 16 GPIOs, PIOOutH and PIOOutL for setting the pin output when the pin is configured as output, PIOUpDownH and PIOUpDownL for selecting the pull-up or pull-down resistor of each pin, PIODriveStrengthH and PIODriveStrengthL for configuring the driving strength of the pin that is configured as an output. In addition, the two read-only registers PIOInH and PIOInL record the status of the 16 GPIOs, and they can be read back by the UPL.

In the test, the GPIO pins used for monitoring should be configured as inputs while the six GPIO pins used for the MUX64 input channel selection should be configured as outputs. By precisely configuring the corresponding registers, all the monitoring paths are read one

by one and their values are checked. Still there is no issue found during the test, the whole monitoring network has a very good connectivity and sufficient ability to play its role in the modular PEB system.

The gain and offset of the lpGBT ADC circuit are also tested under various ambient temperatures. The ADC has four configurable gains (2x, 8x, 16x and 32x), each of which is successively evaluated during the test. The idea is to measure the ADC response curve at each specific temperature point by gradually adjusting the climate chamber working temperature, and the gain and offset of the ADC can be extracted from the curve. The result shows that the two parameters vary with not only temperature but also individual variation of each lpGBT chip. Therefore, for the PEB design, it's necessary to reserve the interface used for injecting external voltage to conduct ADC calibration of each lpGBT. Another possible way of conducting ADC calibration is to directly use the related calibration data that is planned to be acquired at the PEB working temperature ($-30^{\circ}C$) from the lpGBT group. In addition, the response of the three temperature sensors (one $10\ k\Omega$ NTC sensor on the modular PEB board, two internal sensors in the lpGBT and VTRx+) are also tested by selecting their outputs as the input signals of the ADC logic.

6.8 Test of clock and fast command distribution

The clock distribution of the modular PEB is mainly carried out by using the timing lpGBT. The clocks of other components are provided by the timing lpGBT E-clocks and phase shifter clocks. Figure 5.8 shows the block diagram of the clock network of the modular PEB system. The timing lpGBT has up to 29 E-clocks and four phase shifter clocks. Further, 14 of the 29 clocks are delivered to the 14 module emulators while one of the phase shifter clocks is fed into the luminosity lpGBT as its reference clock. This test just aims to check the configuration and connectivity of the E-clocks since the phase shifter reference clock is already validated in the previous tests, such as the test of lpGBT cascade. In terms of the experiment setup, the full-equipped modular PEB joint test system is used. There is no special requirement on the firmware, so we just used the same firmware with the full data path test system.

For the lpGBT configuration, it uses the same configuration files with the full data path test system except that the registers about the E-clocks configuration are set in a different way. Each E-clock has two dedicated registers EPCLKChnCntrH and EPCLK0ChnCntrL used for configuring its different features including the driving strength, inversion, frequency and pre-emphasis. In this test, each of the 14 E-clocks routed to the module emulators is set with different frequencies in sequence. For every frequency, the E-clock is measured at the end of the clock distribution path, which is located on the module emulator board. More specifically, the oscilloscope probe is placed on the specific pin of the 71-pin FH26W

connector. Finally, based on the waveform, the performance of the clock can be deduced. After thorough tests, it's found that all the 14 E-clocks can successfully reach their destination pins on module emulators with correct frequencies after sequentially passing through the lpGBT daughter board, carrier board, FH26W daughter board and flat cable, which means the successful E-clock configuration and a good connectivity of the modular PEB clock network.

The pre-emphasis parameter scan of the lpGBT downlink E-link (fast command) is performed. There are four related parameters that need to be scanned, which are the driving strength varying from 0 to 7, pre-emphasis strength varying from 0 to 7, pre-emphasis mode varying from 0 to 3 and width of pre-emphasis pulse also varying from 0 to 7 respectively for determining the amplitude of output pulse, height, strategy and width of the pre-emphasis pulse. The output signal from each downlink E-link is a $40\ MHz$ clock, which is measured by an oscilloscope. Based on the performance of the observed waveform, a set of optimal values for the four parameters are determined and they are respectively 7, 7, 3 and 0.

The jitter of the lpGBT E-clock is also measured in many aspects. The strategy is to firstly configure the E-clock with a specific frequency ($40\ MHz$, $320\ MHz$ or $640\ MHz$), then measure the clock signal at different positions along its transmission lines with a high-bandwidth differential probe of oscilloscope. The measurement points are either on the modular PEB carrier board or on the module side that is connected to the carrier board through a $70\ cm$ FLEX tail. All the 14 E-clocks that are routed to modules are tested. The result shows that the measured jitters are always less than $9\ ps$ and they are smaller when pre-emphasis is enabled. The skew between the E-clock and fast command is also measured by comparing their waveforms. All the 14 fast command lines from lpGBT downlink E-ports are also tested. It's found that the skews are always $1.5\ ns$.

6.9 Summary

6.9.1 Review of the modular PEB design and test

The modular PEB system has gone through the individual test and joint test where many aspects of the system are tested and evaluated. We can credibly say that modular PEB system fulfills the requirements raised before its design and it is completely functional from the perspective of hardware design. Therefore, it is already qualified to be used in other tasks as a testing and verification platform in the future. From the design and test of this platform, those pending issues and concerns we had about the future PEB design can also be settled.

First of all, the versatile links (lpGBT + VTRx+) are implemented in the modular PEB in the form of the combination of two lpGBT daughter boards and one VTRx+ daughter

board on the carrier board, which leads to two versatile uplinks and one downlink that have already been tested and validated in Section 6.5.1. The successful implementation and operation of the versatile links are carried out and their reliability and robustness on data transmission reassure us about the versatile link application on PEB design. Secondly, the lpGBT cascade strategy with the timing lpGBT as master and the luminosity lpGBT as slave is also implemented in the modular PEB. The connectivities between the two lpGBTs such as the connections of EC ports, reference clock lines, I^2C bus and reset line, are tested and working as expected. Since there is no issue on the hardware aspect for the lpGBT cascade mode, more advanced tests like the EC communication test and I^2C communication test can be performed in the future. We can say that a robust lpGBT cascade structure is already provided by the modular PEB and it can be further tested to help us to build more confidence on dealing with the lpGBT cascade. What happens afterwards would not require too much extra effort.

Thirdly, one of the biggest gains from the modular PEB design and test is that the lpGBT configuration has been successfully executed in an extensive way aided by the dedicated UPL toolkit. Many individual tests and joint tests can all demonstrate the robustness of the lpGBT configuration. We also have better understanding of the lpGBT ASIC and accumulate more and more expertise about its usage and operation, which will definitely benefit not only the configuration of this modular PEB system but also the design of the PEB boards. To resolve the concerns on the complicated PEB design, mastering the configuration and operation of the most intricate ASIC lpGBT is an important step forward. From the perspective of problem solving, it really hits the nail on the head. Although the lpGBT configuration is just performed through the I^2C communication, the other ways of configuration like EC and IC communications are actually equivalent to the I^2C communication since they are all interacting with the same register array. In addition, the other ways are already checked fine in connectivity, and they can be easily tested when the FELEX (FrontEnd LInk eXchange) Phase-II card is used as the DAQ board. It's worth mentioning that the FELEX Phase-II card is the main component developed for the ATLAS readout system in HL-LHC.

Fourthly, the special connectivity method between lpGBT E-ports and module emulator data links is implemented in the modular PEB, so that many data rate patterns can be configured and applied on the lpGBT E-ports. In the design of PEB boards, there must be many data rate patterns used and each data rate pattern requires the hardwired connection between lpGBT E-ports and module emulator data links, therefore those data rate patterns have to be verified in advance. This verification can be easily carried out with the modular PEB system as described in Section 6.5.3. Without the modular PEB, it will be very difficult to conduct the data rate pattern test. Thanks to the special connectivity method, tests of many data rate patterns are supported by the modular PEB, otherwise many hardware

versions corresponding to different data rate patterns have to be designed. Fifthly, the interfaces to modules (or emulators) are implemented in the modular PEB by adopting the exactly the same 71-pin and 13-pin FH26W connectors with the real modules, so that 14 module emulators or real modules are supported. In the tests performed so far, 14 module emulators are attached to the modular PEB as the data sources of the two lpGBTs. In the future, when the HGTD modules are ready, they can be easily integrated into the system. It is very promising that the modular PEB is adjusted as a module batch test platform where 14 modules can be tested at the same time. Sixthly, the modularity and thus compatibility with commercial ASICs are both successfully implemented in the modular PEB as well, which significantly benefits the tests that we performed, especially the individual tests where the separated daughter board is easily tested by simply inserting into the DAQ board.

Seventhly, the POFV structure is implemented on the lpGBT daughter board design. From those intensive tests involving this board, there is no related issue and bug encountered, which means the POFV technology doesn't have any adverse effect on the performance of the board. Therefore, it can be used in the design of PEB board. This is really a good news for the component layout and wire-routing of the PEB design. Lastly, the distance between two neighboring FH26W connectors is evaluated. The adopted distance on the board is $10\ mm$. From the operations of the flat cable attachment and detachment to/from the FH26W connectors, it's found that the distance can be further reduced for the PEB board design, which will also alleviate the congestion of the board surface. An optimal distance can be determined by some more tests. From the tests performed so far, the result is already quite promising.

6.9.2 Serve as module test platform and beyond

After the extensive and thorough tests on the modular PEB system, the HGTD modules can be tested with the help of this system serving as a module batch test platform. The modular PEB system is capable of handling the test of up to 14 modules at the same time, which can significantly accelerate the speed of module test compared with the currently-used test platform where only two ALTIROCs can be tested at a time. The modular PEB system is a good candidate for the future test of the final version of modules before installation.

To adjust the system to support module test, only extra software development work is needed since the hardware is already verified and completely ready. It should be noted that the uFC is no longer used as the DAQ board of the system ever since the FELIX system is setup. The software is actually a script template dedicated for the modular PEB system and it serves as the bridge between hardware and users. There are four main functions that the software should realize, which are the hardware-related initialization for making the modular PEB in an operational state, module initialization and configuration for setting the modules under test as requested by user's specific module test requirement, ADC calibration for

activating the monitoring ability of the system, module readout for performing each test item required by the module test. The software is developed with nested structure where higher-level class object invokes lower-level class object until the basic class object that is instantiated from the serial communication class provided by the FELIX project. More specifically, all the specific operations are wrapped into higher-level classes. For example, the class modular_PEB is used to operate the components on modular PEB such as VTRx+, timing lpGBT, luminosity lpGBT and MUX64, and it invokes the lpGBT class developed by the lpGBT design group. The class Altiroc2 is used to configure the ALTIROC ASICs on front-end modules and it invokes the classes from FADA package developed by the ALTIROC design group. In addition, the configuration file for lpGBT registers is synchronized with the FELIX configuration file by a dedicated script, which can guarantee the correctness of the configuration file used by the script template.

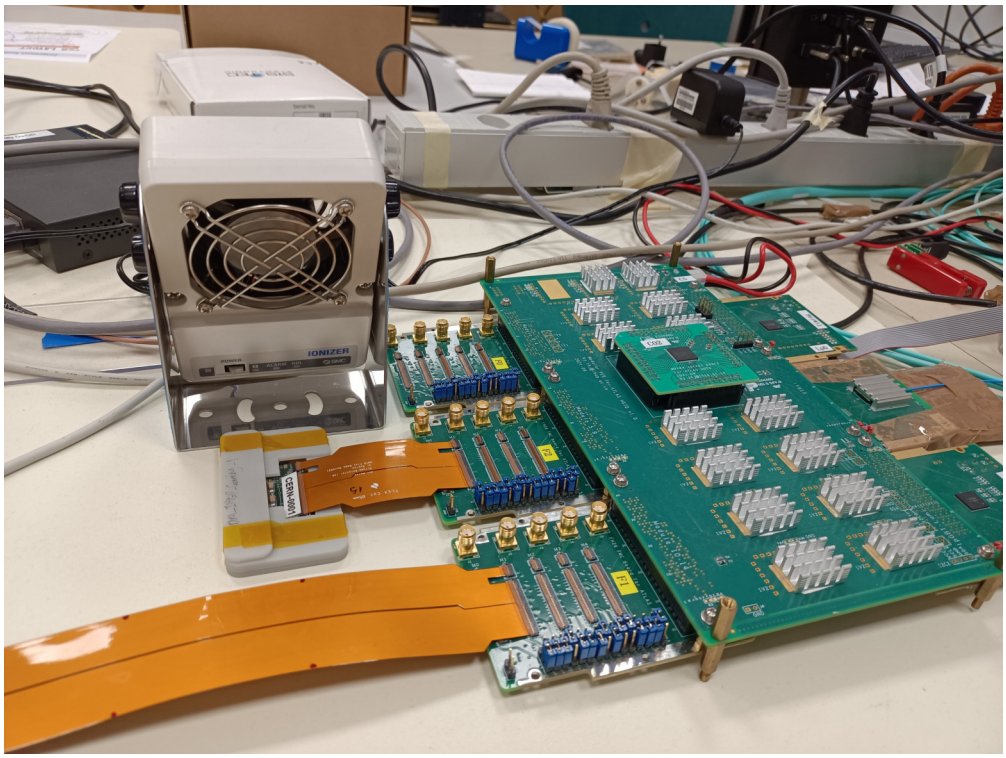


Figure 6.22: Modular PEB system setup at the lab in CERN building 180.

Afterwards, with the implementations of ALTIROC fast command and 8b10b decode in FELIX firmware, the entire modular PEB system is completely functional for module test and normal ALTIROC data taking such as threshold scan and charge scan can be carried out. Then in addition to IHEP, the modular PEB system are also built up at other sites like CERN, Nikhef and KTH for a variety of purposes including not only the module test but also the HGTD DAQ system development, demonstration system establishment, and etc. Figure 6.22 shows the setup at the lab in CERN building 180. It has been seen that the system is already playing an important role in many related tasks and it will keep helping us to conduct more tests and verifications.

Chapter 7

Conclusion and Outlook

In this thesis, two main works are presented at length. One is the physics analysis, i.e. legacy search for the non-resonant production of Higgs boson pairs via ggF and VBF modes in the $b\bar{b}\tau^+\tau^-$ final state, using the full Run-2 proton-proton collision dataset with an integrated luminosity of 140 fb^{-1} collected by the ATLAS detector at the LHC. It is actually the second round of the analysis primarily targeting at a more strict constraint on the Higgs boson self-coupling strength. The other one is hardware work, i.e. design and test of modular PEB, serving as an intermediate product before the real PEB and a verification platform for determining many pending issues about the PEB design. It belongs to the research and development of HGTD peripheral electronics for the ATLAS Phase-2 upgrade during the future HL-LHC period.

In terms of the $b\bar{b}\tau^+\tau^-$ analysis, it is actually a refined research based on the previous round. Even still using the same dataset, the analysis sensitivity gets improved by 10% to 20% for different parameters due to some main updates. For example, in order to increase the sensitivity to κ_λ and κ_{2V}, the categorization is optimized such that finer ggF SRs and dedicated VBF SRs are defined. Also the BDT strategy is optimized and improved for constructing the discriminant variable, which enhances the sensitivity to SM HH and other processes with anomalous values of κ_λ and κ_{2V}. No significant excess above the expected background from Standard Model processes is observed. The observed 95% CL upper limit on the HH signal strength μ_{HH} is 5.9 while the expected one is 3.3 under the assumption of background-only. The reason why the observed limit is looser than the expected limit is that there is a slight data excess in the the high-mHH region, in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR. The observed (expected) 95% confidence intervals for κ_λ and κ_{2V} are respectively $-3.1 < \kappa_\lambda < 9.0$ $(-2.5 < \kappa_\lambda < 9.3)$ and $-0.5 < \kappa_{2V} < 2.7$ $(-0.2 < \kappa_{2V} < 2.4)$.

The next step for this search is to get the Run-3 data included in the analysis, expecting further improvement on the constraint of the Higgs boson self-coupling strength and quartic coupling of Higgs bosons and vector bosons. Also it's foreseen that the analysis can benefit from the improved performance in the reconstruction and identification of physics objects during Run-3. Meanwhile, the result of this analysis can be combined with the latest results

of other HH analyses, so that the sensitivity will be also raised. Another interesting topic is to consider the HH events with two τ-leptons and one large-R jet merged from two small-R b-jets in this analysis. In such a way, more HH events are exploited, thus its sensitivity is expected to get improved, especially the constraint on the modifier κ_{2V}, which can be improved by about 10% from a dedicated investigation describedin Appendix A.1. In anticipation of future research, we expect the emergence of proof for the production of Higgs boson pairs through SM processes during the HL-LHC. A study based on the combined results of previous round [163], has projected the sensitivity of the $b\bar{b}\tau^+\tau^-$ decay channel to the HL-LHC dataset. The projected significance of detecting an excess of SM HH events over the SM background events is 2.8 standard deviations [1].

In terms of the hardware work, the modular PEB is designed and successively goes through individual and joint tests. It's found that the modular PEB satisfies the requirements put forward in Section 5.1 and works as expected in many key aspects. With the joint test system, the following most important functions of the modular PEB are tested, including system configuration, data transmission, power supply distribution, monitoring, clock and fast command distribution. All the functions can be successfully and robustly performed by the modular PEB system, which means the feasibility of the hardware itself and also the ideas adopted in the design. We can confidently assert that the modular PEB is fully operational in terms of hardware design, making it eligible for utilization in future tasks as a testing and verification platform. With the experience and lessons learned from this intermediate modular PEB, we certainly develop more confidence on the design of PEB board. The settlement of many pending issues and concerns further reassures us that a well-functioning and resilient design of HGTD peripheral electronics can be achieved.

In addition to the hardware design of the modular PEB, many auxiliary hardwares, such as the module emulators and the UPL board for lpGBT configuration, are also developed to be used in the joint test system. To implement desired functions on the module emulators, lpGBT emulators and DAQ board, dedicated firmwares are developed to respectively mimic the digital function of ALTIROC, mimic the digital function of lpGBT and serve as the DAQ counterpart of lpGBT. Those auxiliary hardwares and firmwares also go through individual tests and finally integrated into the joint test system. They play an important role on the successful implementation of the individual and joint tests.

Looking ahead to the future tasks that the modular PEB can involve, first of all, in the near future, it can be easily adjusted to a batch test system for modules. The large amount of HGTD modules need to be tested in the perspective of functionality after production, so this adapted test platform supporting up to fourteen modules at the same time can greatly accelerate the module tests and make our life easier. The required adjustments are only from

[1] The numbers have been calculated by halving the b-tagging and theoretical uncertainties, while ignoring the τ_{had} uncertainties stemming from data statistics.

the firmware and software sides since the hardware is already there and tested, therefore the expected time for the adjustments would be very little. It can also be used for the HGTD DAQ system development before the real PEB board is available. It is a quite lightweight hardware version of PEB, but it is convenient and efficient for this task. There are already some groups that use the modular PEB for this purpose. Another application is to be used in the HGTD demonstration system. It has already been used for this purpose at the lab in CERN building 180, where many tests are already performed and more studies will be carried out. In the future, many dedicated tests related to PEB can all be performed with the help of the modular PEB system, especially some tests that may cause damages to the hardware, such as irradiation test, thermal test, power consumption test, HV test and mechanical test on connectors and FLEX cables. It's expected that the modular PEB system can always plays a role in different tasks before the installation of HGTD.

Bibliography

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1.

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30.

[3] J. Alison et al., *Higgs boson potential at colliders: Status and perspectives*, Rev. Phys. **5** (2020) 100045, ed. by Biagio Di Micco, Maxime Gouzevitch, Javier Mazzitelli, and Caterina Vernieri.

[4] J. Baglio et al., *The measurement of the Higgs self-coupling at the LHC: theoretical status*, Journal of High Energy Physics **2013** (2013) 151.

[5] Fady Bishara, Roberto Contino, and Juan Rojo, *Higgs pair production in vector-boson fusion at theC and beyond*, The European Physical Journal C **77** (2017) 481.

[6] Jason Veatch, *Searches for Resonant Scalar Boson Pair Production Using Run 2 LHC Proton-Proton Collision Data*, Symmetry **14** (2022) 260.

[7] ATLAS Collaboration, *Combination of searches for non-resonant and resonant Higgs boson pair production in the $b\bar{b}\gamma\gamma$, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}b\bar{b}$ decay channels using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2021-052, 2021.

[8] Sheldon L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579.

[9] A. Salam and J.C. Ward, *Electromagnetic and weak interactions*, Physics Letters **13** (1964) 168.

[10] Steven Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264.

[11] Peter W. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12** (1964) 132.

[12] Peter W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508.

[13] Peter W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (4 1966) 1156.

[14] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321.

[15] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585.

[16] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (5 1967) 1554.

[17] Particle Data Group, *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020** (2020), 083C01.

[18] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, **2/2017** (2016).

[19] R. Frederix et al., *Higgs pair production at the LHC with NLO and parton-shower effects*, Phys. Lett. B **732** (2014) 142.

[20] J. Baglio et al., $gg \to HH$ *: Combined uncertainties*, Phys. Rev. D **103** (2021) 056002.

[21] Frédéric A. Dreyer and Alexander Karlberg, *Vector-Boson Fusion Higgs Pair Production at* N^3LO, Phys. Rev. D **98** (2018) 114016.

[22] Fabio Monti et al., *Modelling of the single-Higgs simplified template cross-sections (STXS 1.2) for the determination of the Higgs boson trilinear self-coupling*, tech. rep., CERN, 2022.

[23] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043.

[24] S. Borowka et al., *Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass Dependence*, Phys. Rev. Lett. **117** (2016) 012001, [Erratum: Phys.Rev.Lett. 117, 079901 (2016)].

[25] ATLAS Collaboration, *Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector*, ATL-PHYS-PUB-2019-007, 2019.

[26] 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*, JHEP **07** (2014) 079.

[27] Rikkert Frederix and Stefano Frixione, *Merging meets matching in MC@NLO*, JHEP **12** (2012) 061.

[28] Dario Buttazzo et al., *Investigating the near-criticality of the Higgs boson*, JHEP **12** (2013) 089.

[29] Giuseppe Degrassi et al., *Higgs mass and vacuum stability in the Standard Model at NNLO*, JHEP **08** (2012) 098.

[30] So Chigusa, Takeo Moroi, and Yutaro Shoji, *State-of-the-Art Calculation of the Decay Rate of Electroweak Vacuum in the Standard Model*, Phys. Rev. Lett. **119** (2017) 211801.

[31] David E. Morrissey and Michael J. Ramsey-Musolf, *Electroweak baryogenesis*, New J. Phys. **14** (2012) 125003.

[32] A. I. Bochkarev and M. E. Shaposhnikov, *Electroweak Production of Baryon Asymmetry and Upper Bounds on the Higgs and Top Masses*, Mod. Phys. Lett. A **2** (1987) 417.

[33] K. Kajantie, M. Laine, K. Rummukainen, and M. Shaposhnikov, *Generic rules for high temperature dimensional reduction and their application to the standard model*, Nuclear Physics B **458** (1996) 90.

[34] Manuel Reichert et al., *Probing baryogenesis through the Higgs boson self-coupling*, Phys. Rev. D **97** (2018) 075008.

[35] Alexandra Carvalho et al., *Analytical parametrization and shape classification of anomalous HH production in the EFT approach*, (2016).

[36] G. Buchalla, M. Capozi, A. Celis, G. Heinrich, and L. Scyboz, *Higgs boson pair production in non-linear Effective Field Theory with full m_t-dependence at NLO QCD*, JHEP **09** (2018) 057.

[37] Oliver Sim Brüning et al., *LHC Design Report v.1 the LHC Main Ring*, CERN Yellow Reports: Monographs (2004).

[38] Oliver Sim Brüning et al., *LHC Design Report v.2 the LHC Infrastructure and General Services*, CERN Yellow Reports: Monographs (2004).

[39] Michael Benedikt, Paul Collier, V Mertens, John Poole, and Karlheinz Schindl, *LHC Design Report v.3 the LHC Injector Chain*, CERN Yellow Reports: Monographs (2004).

[40] ATLAS Collaboration, *ATLAS: Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN*, 1994.

[41] CMS Collaboration, *CMS, the Compact Muon Solenoid: technical proposal*, LHC technical proposal, Geneva: CERN, 1994.

[42] LHCb Collaboration, *LHCb : Technical Proposal*, Geneva: CERN, 1998.

[43] ALICE Collaboration, *ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC*, LHC technical proposal, Geneva: CERN, 1995.

[44] CERN, *Accelerator complex*, https://home.cern/resources/faqs/facts-and-figures-about-lhc.

[45] CERN, *Dessin representant le detecteur ATLAS*, https://cds.cern.ch/record/39038, 1998.

[46] TikZ.net, *ATLAS coordinate system*, https://tikz.net/category/physics/particle-physics/.

[47] M Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, 2010.

[48] ATLAS collaboration, *ATLAS inner detector*, https://atlas.cern/updates/experiment-briefing/inner-detector-alignment.

[49] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[50] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*, Technical design report. ATLAS, Geneva: CERN, 1996.

[51] CERN, *ATLAS Calorimeter*, https://atlas.cern/Discover/Detector/Calorimeter.

[52] Flickr.com, *ATLAS Muon Spectrometer*, https://www.flickr.com/photos/atlasexperiment/.

[53] Georges Aad et al., *Operation of the ATLAS trigger system in Run 2*, JINST **15** (2020) P10004.

[54] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317.

[55] Rosa Simoniello, *The ATLAS Level-1 Topological Processor: from design to routine usage in Run-2*, (2018).

[56] T Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*, Journal of Physics: Conference Series **119** (2008) 032014.

[57] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444.

[58] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 Data at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-051, 2015.

[59] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, 2012.

[60] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **76** (2016) 292.

[61] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063.

[62] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490.

[63] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466.

[64] ATLAS Collaboration, *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*, (2015).

[65] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 970.

[66] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, (2017).

[67] ATLAS Collaboration, *Evidence for the $H \longrightarrow b\bar{b}$ decay with the ATLAS detector*, JHEP **12** (2017) 024.

[68] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV*, Eur. Phys. J. C **79** (2019) 836.

[69] David Krohn, Jesse Thaler, and Lian-Tao Wang, *Jet trimming*, Journal of High Energy Physics **2010** (2010) 84.

[70] S. Catani, Yu.L. Dokshitzer, M.H. Seymour, and B.R. Webber, *Longitudinally invariant $K(t)$ clustering algorithms for hadron-hadron collisions*, Nuclear Physics B **406** (1993) 187.

[71] Stephen D. Ellis and Davison E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. **D48** (1993) 3160.

[72] ATLAS Collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s}$ = 8 TeV*, Eur. Phys. J. C **76** (2016) 154.

[73] ATLAS Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, ATLAS-CONF-2016-035, 2016.

[74] ATLAS Collaboration, *Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs($\longrightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017.

[75] Matteo Cacciari and Gavin P. Salam, *Pileup subtraction using jet areas*, Physics Letters B **659** (2008) 119.

[76] Matteo Cacciari, Gavin P Salam, and Gregory Soyez, *The catchment area of jets*, Journal of High Energy Physics **2008** (2008) 005.

[77] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016.

[78] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **78** (2018) 903.

[79] ATLAS Collaboration, *Search for resonant and non-resonant Higgs boson pair production in the* $b\bar{b}\tau^+\tau^-$ *decay channel using* 13 *TeV* pp *collision data from the ATLAS detector*, JHEP **07** (2023) 040.

[80] ATLAS Collaboration, *Luminosity determination in* pp *collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **83** (2023) 982.

[81] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823.

[82] S. Agostinelli et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[83] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852.

[84] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016.

[85] NNPDF Collaboration, Richard D. Ball, et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244.

[86] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152.

[87] Enrico Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034.

[88] Andy Buckley et al., *LHAPDF6: parton density access in the LHC precision era*, Eur. Phys. J. C **75** (2015) 132.

[89] Jon Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001.

[90] ATLAS Collaboration, *ATLAS Pythia 8 tunes to* 7 *TeV data*, ATL-PHYS-PUB-2014-021, 2014.

[91] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012.

[92] Massimiliano Grazzini et al., *Higgs boson pair production at NNLO with top quark mass effects*, JHEP **05** (2018) 059.

[93] *LHC Higgs Cross Section HH Sub-group.*

[94] M. Bähr et al., *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639.

[95] Johannes Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196.

[96] The NNPDF Collaboration, Richard D. Ball, et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040.

[97] Frédéric A. Dreyer and Alexander Karlberg, *Vector-boson fusion Higgs pair production at N3LO*, Physical Review D **98** (2018).

[98] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040.

[99] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070.

[100] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043.

[101] NNPDF Collaboration, *Parton distributions for the LHC Run II*, (2014).

[102] Torbjorn Sjostrand et al., *An Introduction to PYTHIA 8.2*, Comput.Phys.Commun. **191** (2015) 159.

[103] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014.

[104] Richard D. Ball et al., *Parton distributions with LHC data*, Nuclear Physics B **867** (2013) 244.

[105] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152.

[106] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015.

[107] Tanju Gleisberg and Stefan Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039.

[108] Federico Buccioni et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866.

[109] Fabio Cascioli, Philipp Maierhöfer, and Stefano Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601.

[110] Ansgar Denner, Stefan Dittmaier, and Lars Hofer, *Collier: A fortran-based complex one-loop library in extended regularizations*, Comput. Phys. Commun. **212** (2017) 220.

[111] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert, *A critical appraisal of NLO+PS matching methods*, JHEP **09** (2012) 049.

[112] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert, *QCD matrix elements + parton showers. The NLO case*, JHEP **04** (2013) 027.

[113] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, *QCD Matrix Elements + Parton Showers*, JHEP **11** (2001) 063.

[114] Stefan Höche, Frank Krauss, Steffen Schumann, and Frank Siegert, *QCD matrix elements and truncated showers*, JHEP **05** (2009) 053.

[115] ATLAS Collaboration, *Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment*, JHEP **08** (2021) 089.

[116] D de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, (2016).

[117] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, JHEP **09** (2014) 145.

[118] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012.

[119] S. Agostinelli et al., *GEANT4 - a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[120] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823.

[121] Mats Joakim Robert Olsson et al., *Search for Higgs boson pair production in $2b + 2l + E_T^{miss}$ final states in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Experiment*, tech. rep., CERN, 2021.

[122] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. A **654** (2011) 481.

[123] Andreas Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007.

[124] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep., New York U., 2012.

[125] Alexander L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693.

[126] ATLAS Collaboration, *Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **843** (2022) 137745.

[127] Andrea Dainese et al., *Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC*, tech. rep., 2019.

[128] ATLAS and CMS Collaborations, *Snowmass White Paper Contribution: Physics with the Phase-2 ATLAS and CMS Detectors*, ATL-PHYS-PUB-2022-018, 2022.

[129] *Projections for measurements of Higgs boson cross sections, branching ratios, coupling parameters and mass with the ATLAS detector at the HL-LHC*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/ PHYS-PUB-2018-054: CERN, 2018.

[130] STEPHEN P. MARTIN, "A SUPERSYMMETRY PRIMER", *Advanced Series on Directions in High Energy Physics*, WORLD SCIENTIFIC, 1998 1.

[131] Nathaniel Craig, *Naturalness: A Snowmass White Paper*, 2022.

[132] O. Aberle, I Béjar Alonso, O Brüning, P Fessia, and Rossi, *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*, CERN Yellow Reports: Monographs, Geneva: CERN, 2020.

[133] CERN, *The HL-LHC key technologies*, `https://home.cern/science/accelerators/high-luminosity-lhc/technologies`.

[134] CERN, *CMS prepares for Phase II*, `https://cerncourier.com/a/cms-prepares-for-phase-ii`.

[135] CERN, *The HL-LHC project*, `https://hilumilhc.web.cern.ch/content/hl-lhc-project`.

[136] Collaboration ATLAS, *Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment*, tech. rep., Draft version for comments: CERN, 2012.

[137] *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*, tech. rep., CERN, 2017.

[138] *ATLAS Liquid Argon Calorimeter Phase-II Upgrade: Technical Design Report*, tech. rep., CERN, 2017.

[139] *Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter*, tech. rep., CERN, 2017.

[140] *Technical Design Report for the ATLAS Inner Tracker Strip Detector*, tech. rep., CERN, 2017.

[141] *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*, tech. rep., CERN, 2017.

[142] *Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*, tech. rep., CERN, 2020.

[143] G. Pellegrini et al., *Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **765** (2014) 12, HSTD-9 2013 - Proceedings of the 9th International "Hiroshima" Symposium on Development and Application of Semiconductor Tracking Detectors.

[144] CERN, *The lpGBT project*, `https://espace.cern.ch/GBT-Project/LpGBT/default.aspx`.

[145] CERN, *The lpGBT manual*, `https://lpgbt.web.cern.ch/lpgbt/v1`.

[146] CERN, *The CERN DCDC project*, `https://espace.cern.ch/project-DCDC-new`.

[147] CERN, *The versatile-link-plus project*, `https://espace.cern.ch/project-Versatile-Link-Plus/SitePages/Home.aspx`.

[148] J. Anderson et al., *FELIX: a PCIe based high-throughput approach for interfacing front-end and trigger electronics in the ATLAS Upgrade framework*, Journal of Instrumentation **11** (2016) C12023.

[149] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, Journal of Instrumentation **13** (2018) P07017.

[150] CERN, *The piGBT manual*, `https://pigbt.web.cern.ch/manual`.

[151] L. Han et al., *Demonstration system of the HGTD peripheral electronics boards for ATLAS phase II upgrade*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1045** (2023) 167651.

[152] ANSI/VITA, *VITA 57.1 standard*, `https://fmchub.github.io/appendix/VITA57_FMC_HPC_LPC_SIGNALS_AND_PINOUT.html`.

[153] Intel, *FRU Information Storage Definition*, `https://www.intel.com/content/dam/www/public/us/en/documents/specification-updates/ipmi-platform-mgt-fru-info-storage-def-v1-0-rev-1-3-spec-update.pdf`.

[154] Xilinx, *7 Series FPGAs Data Sheet: Overview*, `https://docs.xilinx.com/v/u/en-US/ds180_7Series_Overview`.

[155] Xilinx, *7 Series FPGAs Packaging and Pinout*, `https://docs.xilinx.com/v/u/en-US/ug475_7Series_Pkg_Pinout`.

[156] Xilinx, *7 Series FPGAs SelectIO Resources*, `https://docs.xilinx.com/v/u/en-US/ug471_7Series_SelectIO`.

[157] Xilinx, *7 Series FPGAs Configuration*, `https://docs.xilinx.com/v/u/en-US/ug470_7Series_Config`.

[158] Xilinx, *DC and AC Switching Characteristics*, `https://docs.xilinx.com/r/en-US/ds189-spartan-7-data-sheet`.

[159] Xilinx, *LVDS Source Synchronous DDR Deserialization*, `https://docs.xilinx.com/v/u/en-US/xapp1017-lvds-ddr-deserial`.

[160] Xilinx, *Clocking Wizard v6.0 LogiCORE IP Product Guide*, `https://docs.xilinx.com/r/en-US/pg065-clk-wiz/Clocking-Wizard-v6.0-LogiCORE-IP-Product-Guide`.

[161] L. Han et al., *The isolated USB programmer board for lpGBT configuration in ATLAS-HGTD upgrade*, Journal of Instrumentation **17** (2022) C03030.

[162] Jie Zhang, Hangxu Li, and Xiaoshan Jiang, *The MicroTCA Fast Control Board for Generic Control and Data Acquisition Applications in HEP Experiments*, IEEE Transactions on Nuclear Science **66** (2019) 1169.

[163] ATLAS Collaboration, *Projected sensitivity of Higgs boson pair production in the $bb\tau\tau$ final state using proton-proton collisions at HL-LHC with the ATLAS detector*, ATL-PHYS-PUB-2021-044, 2021.

Appendix A

Auxiliary materials on HH search

A.1 Semi-boosted study

A.1.1 Motivation

The primary goal of the semi-boosted study is to exploit more HH events that are uncharted yet by physics analyzers from the ATLAS Run-2 dataset. The study is performed based on the HH decay channel of $b\bar{b}\tau^+\tau^-$, including both the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ sub-channels. The idea is to include those new HH events with one Higgs boson decaying to two -leptons and another Higgs boson decaying to one large-R jet merged from two b–jets into the resolved analysis (where the two -leptons and two b–jets are all resolved). Therefore, those new HH events are called semi-boosted events, thus the related analysis is called semi-boosted analysis. With those more events also involved into the resolved analysis, the whole analysis sensitivity is expected to get increased. This semi-boosted study aims to answer the question of how much the sensitivity gets improved after taking into account the semi-boosted events. As a pioneer study, it is just performed only with statistical uncertainty, trying to give people a ballpark of the gain from semi-boosted events and provide a guidance to the future HH analysis.

Intuitively, it's very likely that there will be more semi-boosted events from the VBF HH process which is caused by much harder proton-proton collisions compared to the ggF process and more likely generates more boosted Higgs bosons that decay to merged b–jets. Figure 1.5(b) shows the leading order Feynman diagrams for the VBF HH process. As can be seen, both the couplings (modifiers) κ_{2V} and κ_λ contribute to the production of this process. Therefore, if there are more semi-boosted events explored, it's expected that the analysis sensitivity on the two coupling modifiers will get improved, especially the sensitivity on the κ_{2V}. Predicted by some Beyond Standard Model (BSM) scenarios, the deviations of the κ_{2V} and κ_λ from their SM values (i.e.1) lead to significant enhancement to the HH production as described in Section 1.2.4, which further motivate the semi-boosted study.

If proved feasible, the inclusion of the semi-boosted events is a quite promising way to

improve the sensitivity on the κ_{2V} measurement, since it is quite cost-efficient and convenient. What the analyzers need to do is to select the HH events with semi-boosted topology from the VBF HH signal samples with varying κ_{2V} and κ_λ, and analyze them together with those resolved HH events. This study will investigate the feasibility of this idea.

A.1.2 Strategy

The general idea of performing this study is to re-use the machinery and strategy of the current HH search in the $b\bar{b}\tau^+\tau^-$ final state as much as possible, including the object reconstruction and identification, trigger selection, background estimation, analysis frameworks and etc. Of course, there are specific requirements and implementations on those aspects dedicated for this semi-boosted study. The study will be performed on both the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ sub-channels. Since the boosted b–jets are not sensitive to the decay mode of the other Higgs boson (either decaying to $\tau_{\text{lep}}\tau_{\text{had}}$ or decaying to $\tau_{\text{had}}\tau_{\text{had}}$), the sensitivity improvements on the two sub-channels should be at the same level.

First of all, for the samples of this study, we can not just re-use the already-existing data and MC samples that are produced for the resolved analysis. Instead, we have to re-produce all the samples to make them suitable for the semi-boosted analysis. The main difference between the samples used for resolved and semi-boosted analysis is the consideration of the large-R jets in the production of the samples. There are large-R jet objects reconstructed in the samples for the semi-boosted analysis while not in the samples for the resolved analysis. For the VBF HH signal samples production, the coupling modifiers κ_{2V} and κ_λ vary from their SM values. Right after the samples production, a truth analysis is performed on the truth level to give us a general view of this study, so that a decision can be made based on the result of the truth analysis. As a result, it's found that the it's very promising to continue this study.

The signal regions (SRs) are defined based on the signal topologies involved in this study. Generally speaking, there are VBF resolved SR and VBF semi-boosted SR, the overlap between the two SRs, and the remaining parts after the subtraction of the overlap region from the two SRs. In terms of the background estimation, the strategy used in the current HH search should perfectly match the semi-boosted study since the samples are in principle the same and they are all targeting at the same phase space. The only issue is the feasibility of the background estimation strategy in the the remaining part after the subtraction of the overlap region from the VBF semi-boosted SR, which is actually a newly-extended phase space where the background estimation strategy should be validated.

For the signal extraction, the multivariate technique is used to train the discriminant for the final simultaneous fit. Generally speaking, in the VBF resolved SR, the BDT trained by the previous round of the analysis [79] is used while may types of discriminants are investigated in the VBF semi-boosted SR. Finally, the statistical analysis is performed by

the commonly-used framework WSMaker in the ATLAS HDBS analysis. For example, with the help of this framework, the likelihood fits can be carried out and the upper limits on the VBF HH production cross section are calculated by successively assuming each of those samples with different κ_{2V} values as the predicted signal of this study. Then the curves of the upper limit on the VBF HH production cross section as a function of the κ_{2V} in different SRs and combination cases can be made as the results of this study. The data points on those curves are a few due to the limited number of signal samples. A linear combination technique that combines some samples with different κ_{2V} as basis samples can parameterize the upper limit as a function of κ_{2V}, so that smooth curves can be obtained.

A.1.3 Samples production

For the samples production, they have to be re-produced to be able to include the large-R jets in the reconstruction. The samples include data samples, signal and background MC samples, and all of them are already in the format of AOD, starting from which they should successively go through the derivation (to DAOD format) and further management (to CxAOD format). For the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the required derivation has been already implemented in the Athena derivation frame by previous studies, and dedicated framework for CxAOD production is also developed before. Therefore, the samples for the $\tau_{\text{had}}\tau_{\text{had}}$ channel are directly produced out of the box after requesting production tasks. For the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, the derivation implementation and CxAOD production framework are newly-adjusted based on the ones used for resolved samples. Similarly, the samples are finally produced by requesting production tasks of the ATLAS simulation.

In terms of the signal samples, those with representative κ_{2V} and κ_{λ} values are produced. For the choosing of the κ_{2V} and κ_{λ} values, it is the same for both the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels. More specifically, four VBF samples with the κ_{λ} values of $0, 1, 2, 10$ and other coupling modifiers fixed to their SM values are produced while six VBF samples with the κ_{2V} values of $0, 0.5, 1, 1.5, 2, 3$ and other coupling modifiers fixed to their SM values are also produced. In terms of the derivation of the samples, the dedicated job option scripts should be implemented in the derivation component of the Athena framework to configure how the objects are processed at this stage and what will be stored in the samples. In terms of the CxAOD production, the dedicated framework called CxAODMaker is adjusted and used for this purpose.

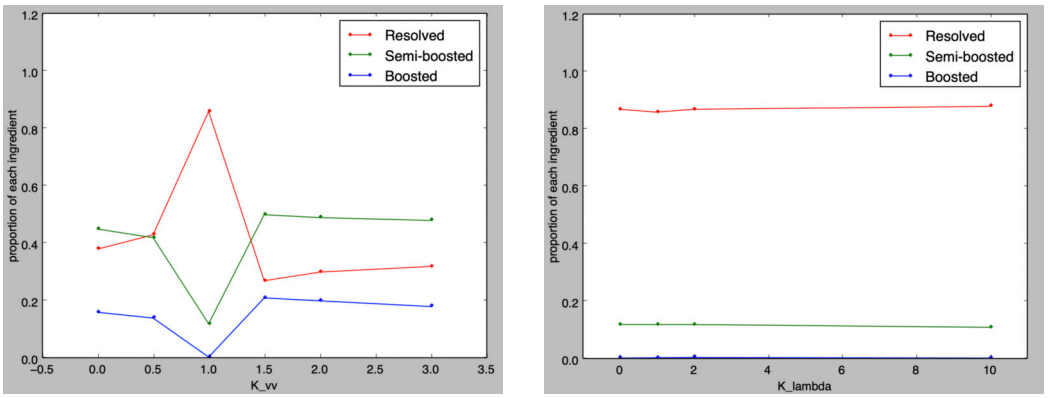
A.1.4 Truth analysis

A truth analysis is firstly performed before the real analysis to check if it's worth doing this semi-boosted study. The idea is to see how many at most semi-boosted events there are in the generated VBF samples with varying κ_{2V} and κ_{λ} as described in Appendix A.1.3.

The VBF samples used for the truth analysis is specifically-produced truth samples, which are different from the generated VBF samples. They are generated based on the event files after parton shower, successively going through the truth derivation that mimics the real derivation and the truth-level selection that also imitates the real analysis selection. Finally, the components of the resolved and semi-boosted events can be estimated and their fractions can also be calculated. It should be noted that the fraction of semi-boosted events is the maximum ratio, in reality, it's less due to the detector acceptance.

The truth selection defines three regions, which are resolved, semi-boosted and boosted regions. Just as their name implies, the regions are enriched with their corresponding types of events. Here the boosted events are the HH events with one large-R jet merged from two -leptons and another large-R jet merged from two b–jets. The threshold cut on the angular distance between the two -leptons $\Delta R(\tau_{\text{had}}, \tau_{\text{had}})$ is 0.4, which means the two -leptons with $\Delta R(\tau_{\text{had}}, \tau_{\text{had}}) < 0.4$ are considered as a large-R jet (boosted). The threshold cut on the angular distance between the two b–jets $\Delta R(b, b)$ is 0.8, which means the two b–jets with $\Delta R(b, b) < 0.8$ are also considered as a large-R jet (boosted).

Figure A.1 shows the results of this truth analysis. On the left plot, it shows the fraction of each type of events in the VBF signal samples with six different κ_{2V} values. As can be seen, there is a considerable amount of semi-boosted events in the VBF signal samples, especially when the κ_{2V} value deviates from its SM value, the amount of semi-boosted events triples or even quadruples. Also at those values deviating from SM value, the semi-boosted events are more than resolved events. On the right plot, it shows the fraction of each type of events in the VBF signal samples with four different κ_λ values. As can be seen, for the semi-boosted events, the numbers at the non-SM values are not enhanced and still at the same level with that of the SM value.



(a) Fraction of each type of events as a function of κ_{2V}

(b) Fraction of each type of events as a function of κ_λ

Figure A.1: Fraction of each type of events (resolved, semi-boosted and boosted) in the VBF samples.

Based on the above-mentioned results, we conclude that the deviation of κ_λ from its SM

value doesn't affect the fraction of semi-boosted events and still keeps it remain at the low level of the SM value, and the deviation of κ_{2V} from its SM value can greatly enhance the fraction of semi-boosted events. Therefore, the four VBF samples with varying κ_λ values will not be used since they can not easily benefit the κ_λ measurement with the inclusion of semi-boosted events; the six VBF samples with varying κ_{2V} values will be used and they will very likely improve the sensitivity on κ_{2V} after the inclusion of semi-boosted events.

A.1.5 Event selection and categorization

The trigger selection and basic event selection of this study follow the previous analysis of HH search [79]. Figure A.2 shows the signal events distribution, which is valid for all the six VBF signal samples with varying κ_{2V} values. Based on the distribution, five different signal regions are defined, which are the resolved SR (Res), Semi-boosted SR (Semi), overlap SR (OL), the remaining region after resolved SR minus overlap (ResMO), and the remaining region after semi-boosted SR minus overlap (SemiMO).

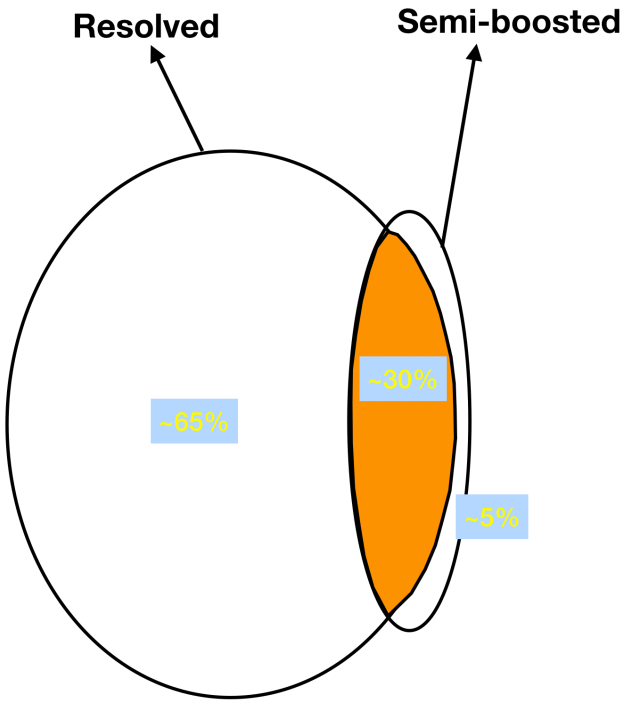


Figure A.2: Signal events distribution that is valid for all the six VBF signal samples with varying κ_{2V} values.

In terms of the definition of the five SRs, they all inherit the basic selection criteria from previous analysis, require the selection of two VBF jets pointing to opposite hemispheres, and demand two -leptons with opposite charges. Then those SRs have their own specific selection criteria. The Semi SR requires one large-R jet with two b–tagged track jets, and a p_{T} cut of $>$ 200 GeV and a $|\eta|$ cut of $<$ 2 on the large-R jet; the Res SR requires two b–tagged jets; the ResMO firstly vetoes one large-R jet with two b–tagged track jets then requires two

b–tagged jets; the SemiMO firstly requires less than two b–tagged jets, then demands one large-R jet with two b–tagged track jets, and a p_{T} cut of > 200 GeV and a $|\eta|$ cut of < 2 on the large-R jet; the OL SR firstly requires two b–tagged jets, then also asks one large-R jet with two b–tagged track jets, and a p_{T} cut of > 200 GeV and a $|\eta|$ cut of < 2 on the large-R jet;

A.1.6 Validation of background estimation strategy used in SemiMO category

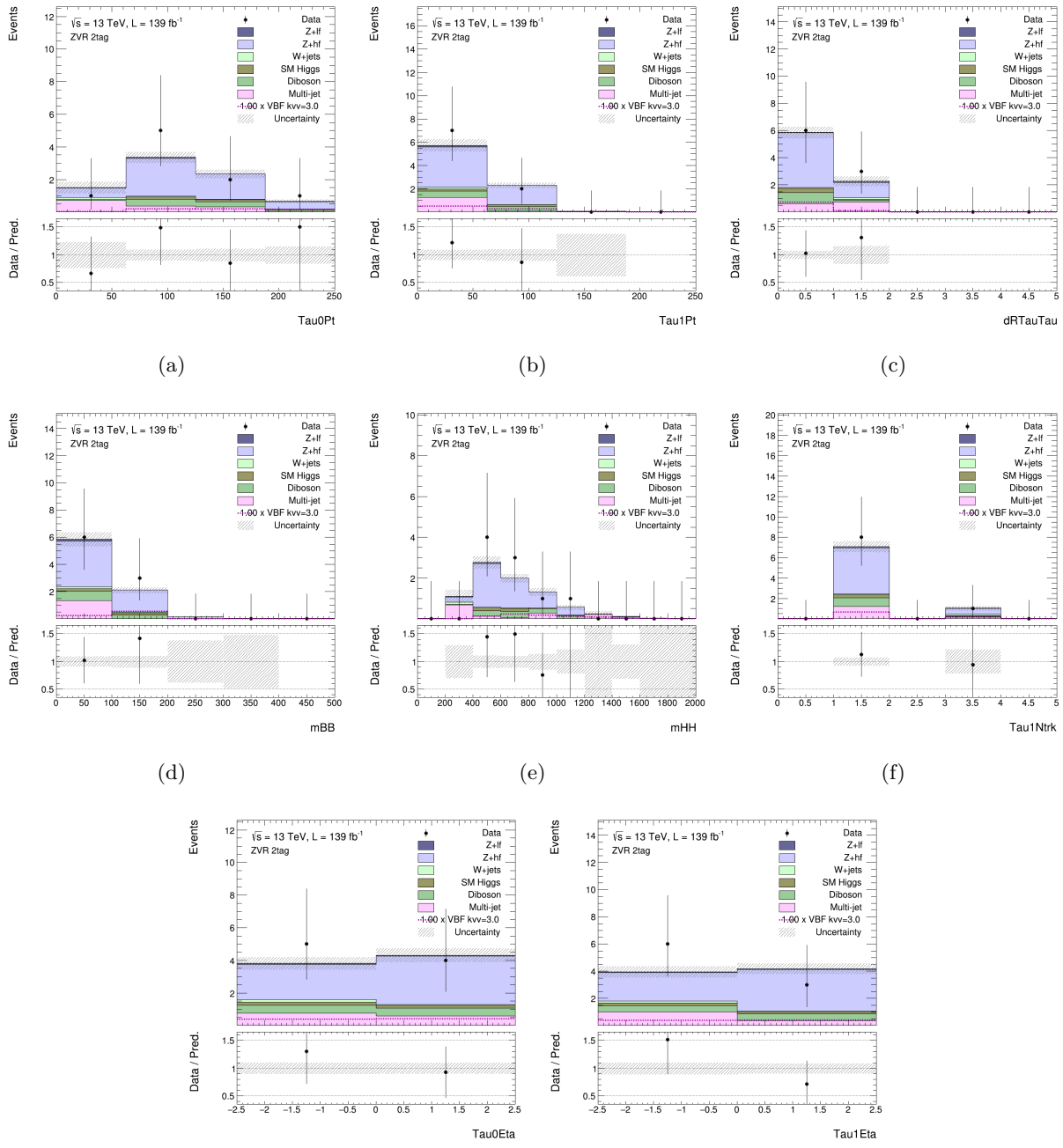
The background estimation strategy of the semi-boosted study follows that of the previous analysis of HH search in the $b\bar{b}\tau^+\tau^-$ final state. In the regime of resolved phase space, the strategy must be valid in this study since there is actually no difference between the two cases. However, in the regime of the SemiMO SR where the previous strategy was not used, it's not guaranteed the strategy can still work. Although it is very likely that the strategy can be used since the estimation of the dominant background Z+HF (accounting for about 70% of the total background) in SemiMO SR is not very sensitive to the topology of the two b–jets in this case. In addition, there are also some components of multi-jet and $t\bar{t}$ with fake-τ_{had}. It's worth performing a validation of the background estimation strategy in the SemiMO SR.

Here the validation in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel is presented as an example. For the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel, it is performed in a similar way. The validation in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel is performed by defining three validation regions, ZVR, FakeVR and TopVR, respectively dedicated for the three main background components (Z+HF, multi-jet and $t\bar{t}$ with fake-τ_{had}) in the SemiMO SR. The definition of the three validation regions all require two -leptons with opposite charges, one or more large-R jets and the leading one with more than one track jets among which there are one or more b–tagged, also demand the selection of two VBF jets pointing to opposite hemispheres. There are specific selections for the three validation regions. The ZVR further requires $70GeV < m_{\tau\tau}^{\mathrm{MMC}} < 110GeV$ and $\Delta R(\tau_{\mathrm{had}}, \tau_{\mathrm{had}}) < 1.8$; the FakeVR requires $m_{\tau\tau}^{\mathrm{MMC}} > 110GeV$ and one b–tagged track jet; the TopVR demands $m_{\tau\tau}^{\mathrm{MMC}} > 150GeV$ and also $m_{bb} > 150GeV$.

The validation plots of ZVR, FakeVR and TopVR are respectively shown in the Figure A.3, Figure A.4 and Figure A.5. As can be seen, the predicted events are quite consistent with the data. Therefore, we conclude that the background estimation strategy is still suitable for the SemiMO SR.

A.1.7 Discriminant variable study

The discriminant variables used in the Res SR and ResMO SR are the already-trained BDTs from previous round of the analysis [79]. They are suitable in the phase space of the

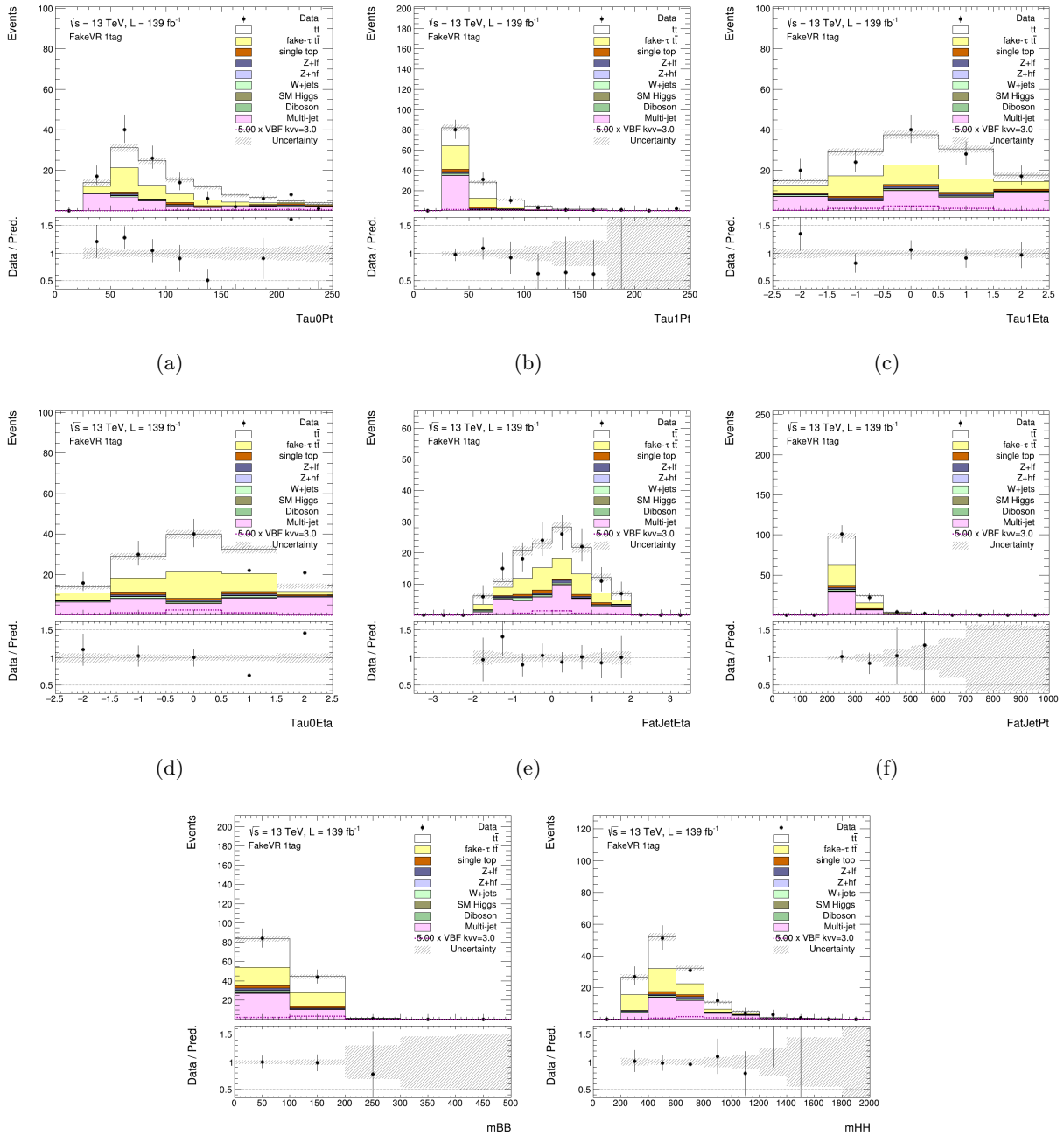


(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure A.3: Validation plots in the ZVR.

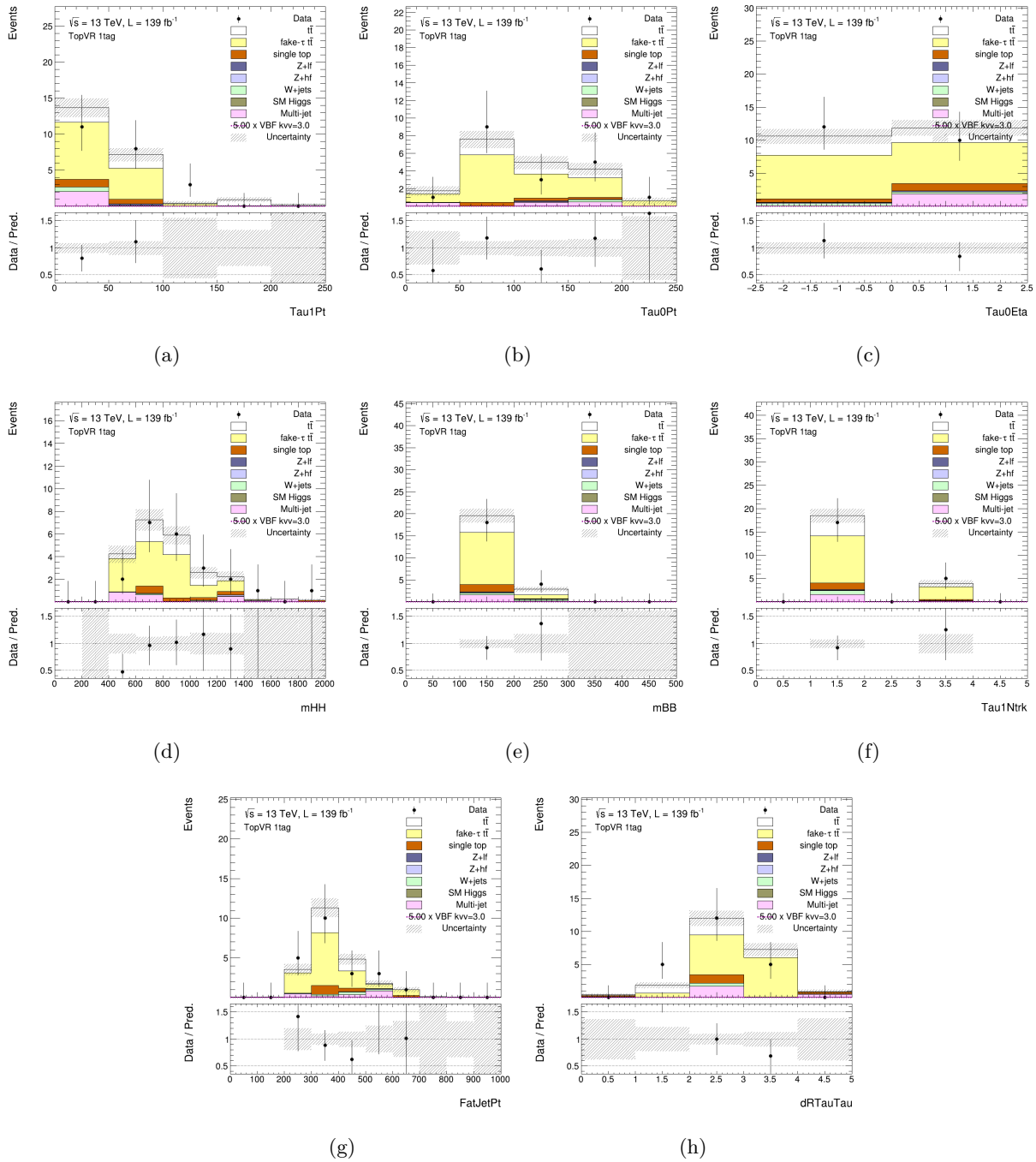


(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure A.4: Validation plots in the FakeVR.



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure A.5: Validation plots in the TopVR.

Res SR and ResMO SR since that's the regime where they were trained before. However, for the discriminant variables used in the Semi, OL and SemiMO SRs where the semi-boosted topology is targeted, new variables need to be created. In this study, two types of discriminant variables are investigated for this purpose, one is a newly-created likelihood ratio while the other is a newly-trained BDT. Both of them are created/trained based on the samples in the Semi SR and applied in in the Semi, OL and SemiMO SRs. Finally, statistical fit will be performed on those SRs with their discriminant variables as the signal extractors.

The likelihood ratio is created from three input variables that already have separation power of signal and background. The VBF signal with $\kappa_{2V} = 3$ is used in the creation of the likelihood ratio. Therefore, it can combine the separation powers of the three variables, which are the invariant mass of the two τ-leptons ($m_{\tau\tau}^{\text{MMC}}$), the invariant mass of the two b–jets (m_{bb}) and the invariant mass of the two VBF jets (m_{jj}^{VBF}). The likelihood ratio is computed with the equation $log(ratio_{m_{\tau\tau}^{\text{MMC}}} \times ratio_{m_{bb}} \times ratio_{m_{jj}^{\text{VBF}}})$, where the $ratio_{m_{\tau\tau}^{\text{MMC}}}$ is the ratio of the fitted probability density function (PDF) from signal to the fitted PDF from background based on the distribution of the $m_{\tau\tau}^{\text{MMC}}$; the $ratio_{m_{bb}}$ and $ratio_{m_{jj}^{\text{VBF}}}$ are calculated in the same way. Figure A.6 shows the distribution of the likelihood ratio in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, which can separate signal from background way better than each individual input variable.

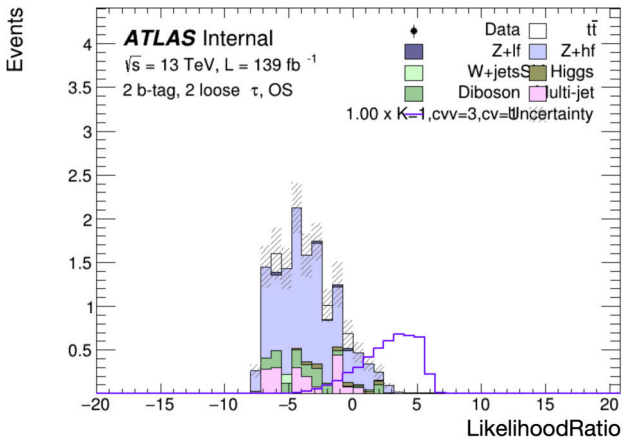


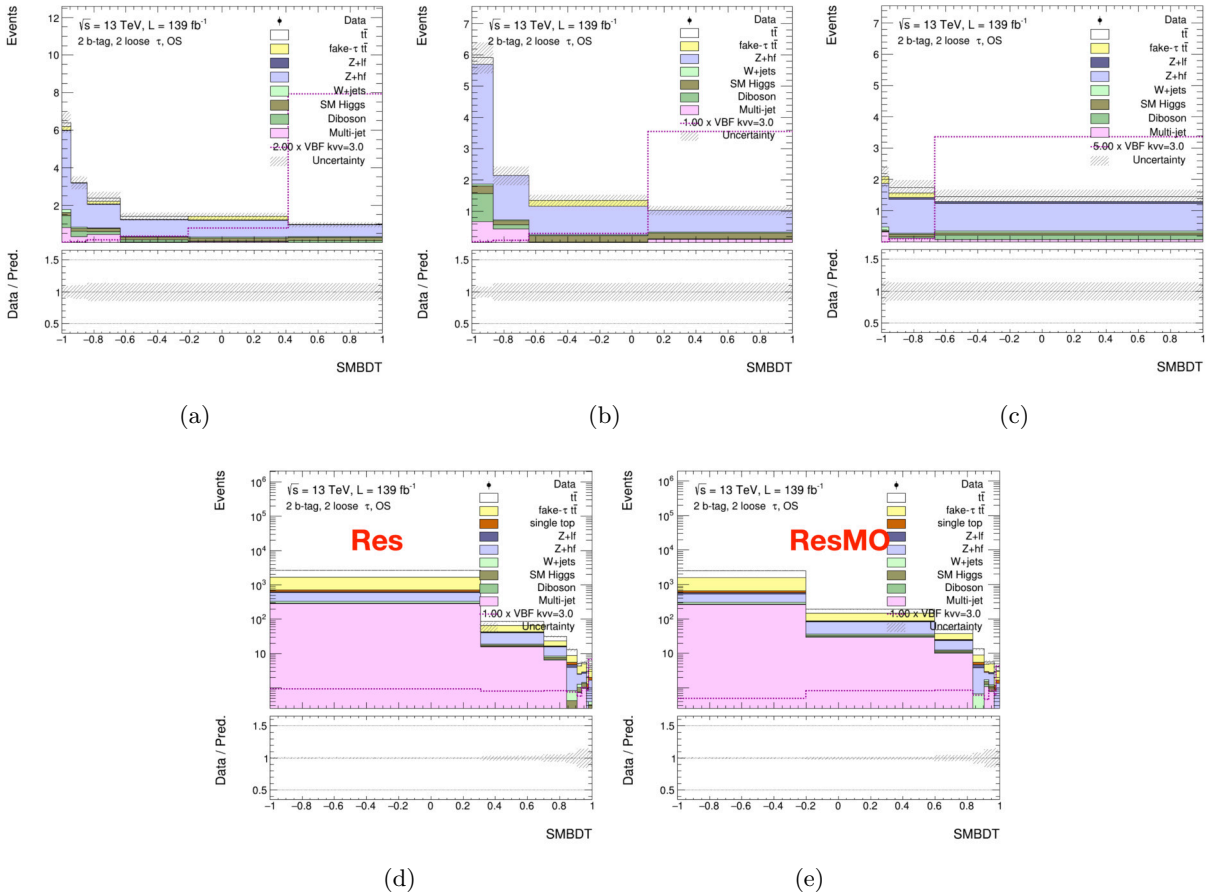
Figure A.6: The distribution of the likelihood ratio as a discriminant variable in the Semi, OL and SemiMO SRs for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The blue line is the VBF signal with $\kappa_{2V} = 3$ while the stacked histograms are the background.

The other type of discriminant variable is a BDT, which is also trained in the Semi SR. The BDT also adopts the VBF sample with $\kappa_{2V} = 3$ as the signal against the combined background. The samples are split into the even and odd parts based on their event numbers, so that the BDT can be trained on one part and applied on the other part. Five input variables are used, which are respectively m_{bb}, $m_{\tau\tau}^{\text{MMC}}$, m_{jj}^{VBF}, m_{HH} and $p_{\text{T}}^{\tau\tau}$. In addition, the hyper-parameters like the number of trees, maximum depth of tree and shrinkage are

optimized.

It was found that the performances of the likelihood ratio and BDT are pretty similar after comparing the significances calculated based on them. Finally, the BDT is chosen as the discriminant variable for the Semi, OL and SemiMO SRs since it is more commonly-used and acceptable by the analysis group.

The discriminant variable distributions in the five types of SRs for the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels are respectively shown in Figure A.7 and Figure A.8. It's worth mentioning that the binning strategy applied on those distributions are similar to the one described in Section 3.6.3, but it requires that the minimum MC statistical uncertainty on the sum of backgrounds of each bin is set to be 10% and the number of expected background events in each bin is required to be greater than one.



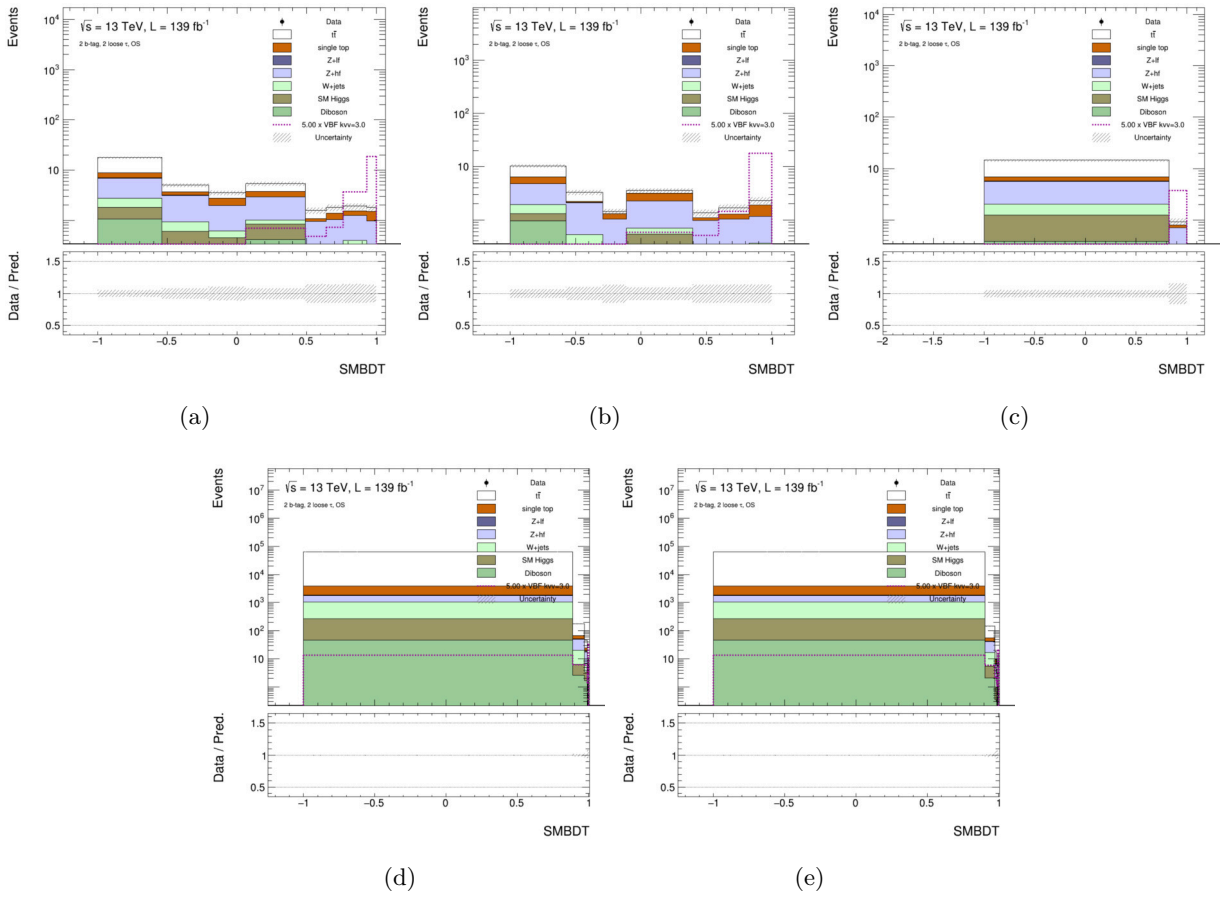
(a) (b) (c)

(d) (e)

Figure A.7: The discriminant variable distributions in the five types of SRs for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The upper three plots are respectively for the (a)Semi, (b)OL and (c)SemiMO SRs while the lower two plots are for the (d)Res and (e)ResMO SRs.

A.1.8 Statistical analysis

The statistical analysis of this semi-boosted study use the same fit model described in Section 3.6.1. Also the same statistical analysis framework called WSMaker is used. For the fit,



(a) (b) (c)

(d) (e)

Figure A.8: The discriminant variable distributions in the five types of SRs for the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. The upper three plots are respectively for the (a)Semi, (b)OL and (c)SemiMO SRs while the lower two plots are for the (d)Res and (e)ResMO SRs.

many different setups can be carried out. For example, the individual fits are performed in each of the five SRs; the ResMO and Semi SRs are fitted simultaneously, so-called Comb1 (i.e. the combination of the ResMO and Semi SRs); the Res and SemiMO SRs are also fitted simultaneously, so-called Comb2 (i.e. the combination of the Res and SemiMO SRs). In each of the fit setups, six statistical fits will be successively performed with the signal strength of each of the six VBF samples with varying κ_{2V} as the POI each time. Thus six upper limits on the signal strength will be obtained from every fit setup. Since the signal strength for a signal sample with specific κ_{2V} value can be converted to the corresponding cross section by knowing the predicted cross section at that certain κ_{2V} point, we can further get six upper limits on the cross section of each of the six VBF samples with varying κ_{2V}. Finally, a curve of upper limits on the cross section as a function of κ_{2V} is created for every fit setup.

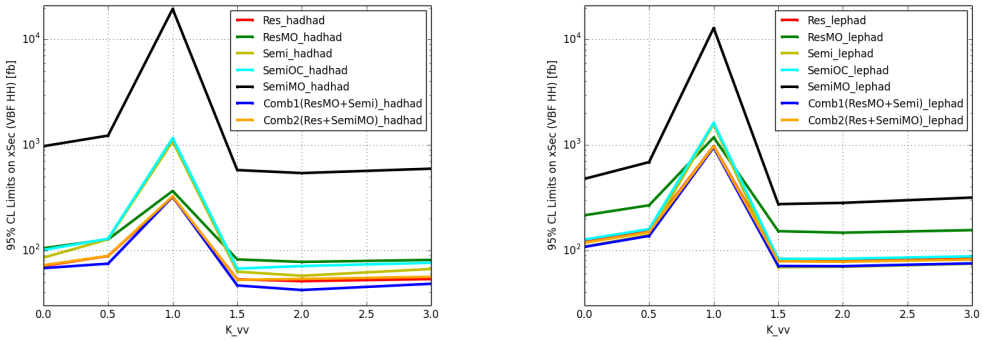
Additionally, an extended upper limit curve can also be made since the above-mentioned curve has just six data points and ranges from 0 to 3 of κ_{2V}. The extended curve is made as follows. Firstly, in each fit setup, instead of using one signal strength of a VBF sample as the POI, it uses six signal strengths of the six VBF samples with varying κ_{2V} as six POIs which are further parameterized as the functions of κ_{2V}, κ_λ and κ_V. Then the six VBF samples are also linearly combined with the same strategy as mentioned in Section 3.1.2.1. Thus a parameterized workspace can be created as a function of κ_{2V}, κ_λ and κ_V for a specific fit setup. When performing the fit, one just need to fix the other two coupling modifiers κ_λ and κ_V to their SM values and vary the target κ_{2V} values as you wish. In this way, the extended upper limit curve can have finer data points and larger κ_{2V} range.

A.1.9 Results

As mentioned previously, the upper limit on the VBF signal cross section is as the figure of merit that implies the sensitivity of this analysis. Figure A.9 shows the upper limits on the VBF signal cross section as a function of κ_{2V} in different SRs and combination cases for both $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels. As can be seen, the upper limit curves for the combination cases always perform better than that for individual SRs due to the consideration of semi-boosted events. It can also be observed that the performance in the case of Comb1 is always better than that in the case of Comb2.

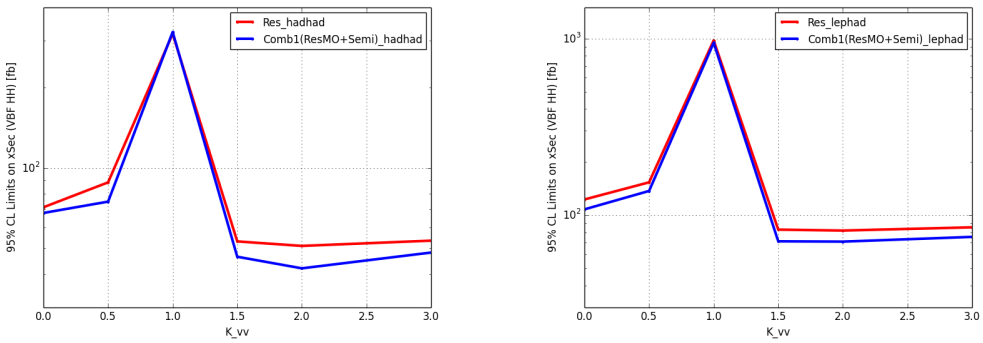
To have a more clear view of how the upper limit gets improved before and after the inclusion of the semi-boosted events, the comparison between Res and Comb1 are made for both channels by just plotting their upper limit curves on the figures. Figure A.10 shows the comparison. We can clearly see the improvements, which are about 8 to 15% for both channels from a quantitative perspective.

Additionally, as mentioned in Appendix A.1.8, the extended upper limit curves are also made to expand the range of κ_{2V} using a linear combination technique of six basis VBF



(a) Upper limit curves for the $\tau_{\text{had}}\tau_{\text{had}}$ channel (b) Upper limit curves for the $\tau_{\text{lep}}\tau_{\text{had}}$ channel

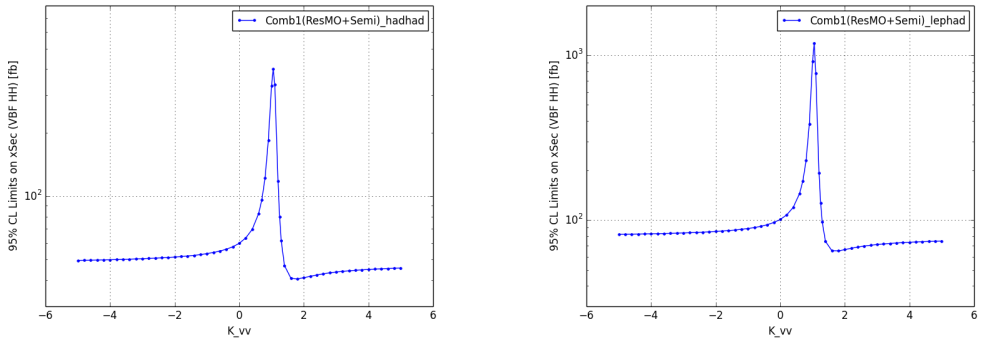
Figure A.9: The upper limit on the VBF signal cross section as a function of κ_{2V} in different SRs and combination cases.



(a) For the $\tau_{\text{had}}\tau_{\text{had}}$ channel (b) For the $\tau_{\text{lep}}\tau_{\text{had}}$ channel

Figure A.10: The upper limit comparison of Res and Comb1 SRs.

samples. Figure A.11 shows the extended upper limit curves of the Comb1 case for both $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels.



(a) For the $\tau_{\text{had}}\tau_{\text{had}}$ channel

(b) For the $\tau_{\text{lep}}\tau_{\text{had}}$ channel

Figure A.11: The extended upper limit curves of the Comb1 case.

A.1.10 Conclusion

From the results presented in Appendix A.1.9, we conclude that introducing extra semi-boosted events in the VBF region by using the VBF signal samples with varying κ_{2V} can improve the sensitivity on κ_{2V} by 8 to 15%, which is already a considerable increase even though the study was performed with just considering the MC statistics. With such an improvement, it's worthwhile to try to include the semi-boosted HH events in the next iteration of the analysis for Run-3. The strategy adopted in this study can hopefully provide some useful thoughts and ideas in the future.

Appendix B

Auxiliary materials on modular PEB

B.1 TPS56428 daughter board

The idea of the TPS56428 daughter board design is to copy the bPOL12V daughter board as much as possible, such as the designs of the stamp hole, shielding case, board size and PCB stack-up. The main difference is the ASIC used on the board. In this design, a commercial DCDC ASIC TPS56428 is used to mimic the functions of the bPOL12V. Figure B.1 shows the top view of the board. As can be seen, the same stamp holes and three shield are located at the same position with bPOL12V daughter board, which allows the board to interface with the carrier board and use the same shielding cases. The ASIC is placed on the bottom-left side of the board and the large inductor is located to its right. Around them, there are those auxiliary components like resistors, capacitors and small inductors.

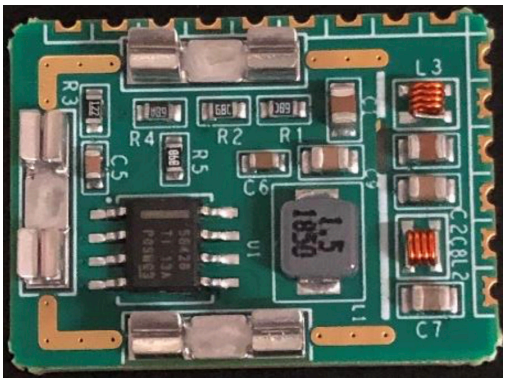


Figure B.1: Top view of the TPS56428 daughter board.

Like the bPOL12V daughter board, all the important pins of the ASIC are also routed to the stamp holes and the pin assignment of those stamp holes is exactly the same. For this board, only the voltage input and output, power enable and power good signals are routed out to the stamp holes, and the stamp holes assigned for temperature sensor and output voltage fine-tuning bits are not connected to any pins since the ASIC doesn't provide those functions. It's required that a 0.1 μF capacitor is used in series between the VBST

and SW pins, the 1st pin is for the supply input for the high-side FET gate drive circuit while the 2nd pin is for the switch node connection between high-side NFET and low-side NFET. The $1.5\ \mu H$ large inductor is used between output voltage and SW pin, and it takes nearly the same space on the board compared to the ASIC. The output voltage is set with a two-resistor (R1 and R2) divider from the output node to the voltage feedback pin VFB, more specifically, the R1 is used between the output node and VFB pin while the R2 pulls VFB pin down to ground. The output value is determined with the formula $0.6 + (1 + R1/R2)$. In the case of $1.2\ V$ output voltage, both R1 and R2 are chosen to be $49.9\ k\Omega$. The filter components including capacitors and inductors are used for the voltage input and output. To achieve appropriate heat dissipation, the exposed thermal pad of the ASIC is soldered to the board and also connected to ground by a large number of vias.

B.2 UPL test

The UPL is designed to configure the lpGBT through its I^2C bus and several dedicated GPIO pins. Therefore, the idea of the UPL test is to check if those pins of the 14-pin header connector can play the roles as expected. As shown in the pin assignment of the 14-pin header connector (**??**), except the GND and NC (not-connected) pins, the performances of all the other pins should be tested.

Figure B.2 shows the setup for the UPL test. As shown in the picture, one lpGBT daughter board is used as the operative subject of the UPL. Since the daughter board is not integrated in a carrier board but used alone, it has to be supplied by an external power supply unit through the reserved external $1.2V$ power supply input connector just for debugging purpose. This power supply can already make sure that most of the lpGBT logics are properly powered on, including the I^2C slave block and register array which are actually the I^2C communication target of the UPL. Because of that, the test can still be properly carried out even without the lpGBT daughter board connected to carrier board. The UPL is connected to the lpGBT daughter board through the flat cable while on the other side it is attached to the host computer through the USB cable.

The most important test item of the UPL is the I^2C communication, which determines the lpGBT register operations thus the configuration of the entire ASIC. Starting from the host computer where the dedicated UPL GUI or its counterpart scripts are installed, the lpGBT register operations (read and write) are firstly initiated by the button clicks on the GUI or script execution from terminal and wrapped into the I^2C data frame, which is transferred out of the host computer in the form of USB protocol through the USB cable to the UPL. Then the UPL extracts the original I^2C data frame out of the USB data and sends out the data in the form of I^2C protocol through its 14-pin header connector. Finally, the SCL and SDA signals reach the lpGBT I^2C slave port and get processed by the related

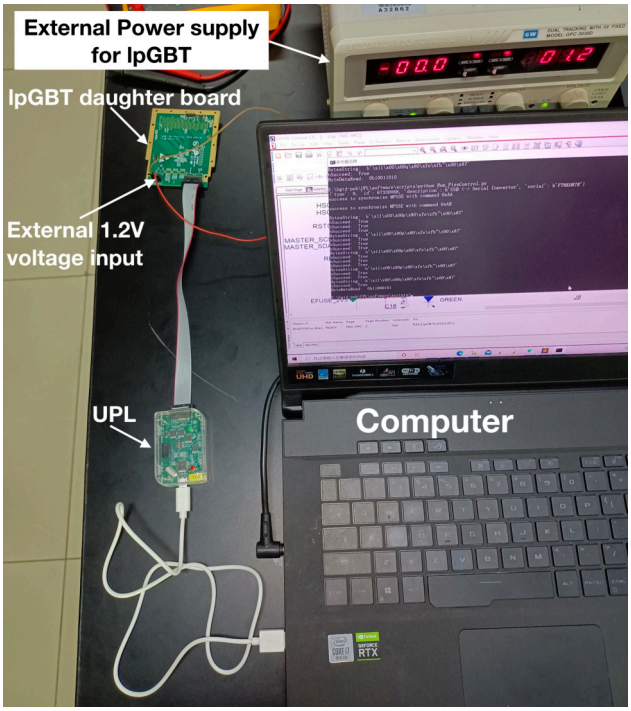


Figure B.2: Picture of the setup for the UPL test.

slave logic, which directly interfaces with the registers based on the communication signals received from the host computer. The communication between the host computer and the lpGBT register is successfully established with the help of the UPL board, which means the main function of the UPL, i.e. I^2C communication, is completely realized, the software and related scripts work as expected.

In addition to the I^2C communication test, there are also some other pins of the UPL connector that need tests. For the pins of MODE0, MODE1, MODE2, MODE3 and RSTB, since they are connected to the UPL GPIOs, their directions can be set to be output and their values are specified through the GUI. On the corresponding points on the lpGBT daughter board, the signals of them can be checked through the probe of multimeter. For the pin of READY indicating the completion of lpGBT configuration, it should always remain low in our setup. Since it is connected to the UPL GPIO, it can be read through the GUI by setting the direction of the GPIO as input. For the pin of 1V2MON indicating the power supply state of lpGBT, it should always remain high. Since it is connected to the UPL GPIO, it can also be read through the GUI by setting the direction of the GPIO as input. For the pin of EFUSE, it is connected to the $2.5\ V$ output of the LDO whose enable signal is also connected to the UPL GPIO, it can be controlled from the host computer through the GUI by setting the GPIO output. Therefore, if the GPIO output is set to be high from the GUI, it's expected that the $2.5\ V$ voltage can be measured by multimeter at the EFUSE pin of the UPL connector. It should be noted that, in order to avoid possible damages to the lpGBT E-fuse array, the lpGBT daughter board should be disconnected from the UPL when the EFUSE pin is tested. In conclusion, after careful tests, all the pins works as expected.

B.3 Test of module emulator and FH26W daughter board

There are quite many test items that the module emulator needs to go through before it can be integrated into the modular PEB board. The test items include the power network test (including static and dynamic test), the connectivity test of the 71-pin FH26W connector and FPGA, FLASH memory and FPGA basic functions. For the test of the FH26W daughter board, only the connectivity of its five 71-pin FH26W connectors and one FMC connector need to be checked. For the connectivity of the 13-pin FH26W connectors on both the module emulator and FH26W daughter board, it's not urgent to check since those connectors are used for HV power supply of LGAD sensors, but it needs to be checked in the future when the real modules are connected to the modular PEB board.

Starting from the power network test of the module emulator, which is simple but essential, the static test is to measure the internal resistance of the different voltages ($1.0\ V$, $1.2\ V$, $1.8\ V$ and $2.5\ V$) to ground before power on to avoid possible damages caused by short-circuit in the power network, and the dynamic test is to measure the values of those different voltages after power on to see if they reach the desired values or not. Both of the tests can be carried out with a multimeter by pointing the positive and negative probes to the certain pads on the module emulator board. Fortunately, all the forty second version module emulators pass the power network test.

In terms of the FLASH memory and FPGA basic functions test, the idea is to firstly program the FLASH memory then load the configuration file (firmware) from the memory to the FPGA, finally check if the logics implemented in FPGA can realize the function as the firmware specifies. Figure B.3 shows the setup of this experiment. In the middle, the module emulator is held by two special clips, the one on the right is used for providing $1.2\ V$ input voltage for the board while the other one on the left is used as the JTAG connector that is occupied by a Xilinx JTAG programmer (HS2 Rev.A). The programmer is further connected to the host computer where the firmware download to module emulator can be initiated. In this experiment, the firmware is quite simple, aiming to regulate the $200\ MHz$ on-board clock to a clock with the frequency of $40\ MHz$ and send it out through a pair of FPGA differential pins which are routed to two test points on the board. Therefore, we just need to measure the $40\ MHz$ clock by putting an oscilloscope probe on the test points and check if the desired waveform is displayed on the screen. The special probe is shown on the bottom-right side of the picture. After the tests, it's found that most of the module emulators work as expected except two of them can not properly load the file from the FLASH memory which is caused by the poor soldering of some related pads.

In terms of the connectivities of the FPGA, 71-pin FH26W connectors on both the module emulator and FH26W daughter board, and the FMC connector on the FH26W

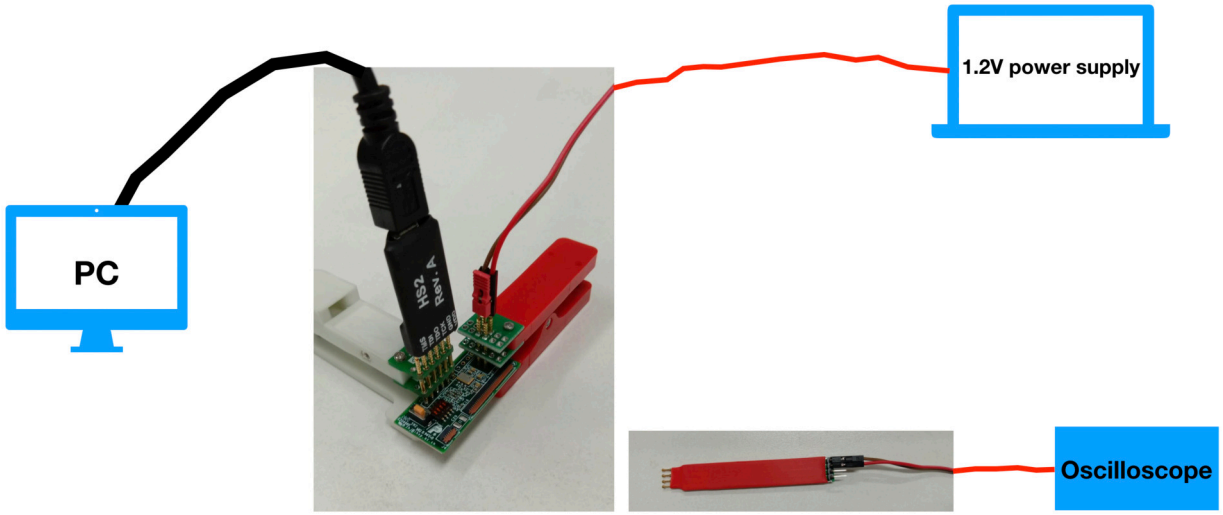
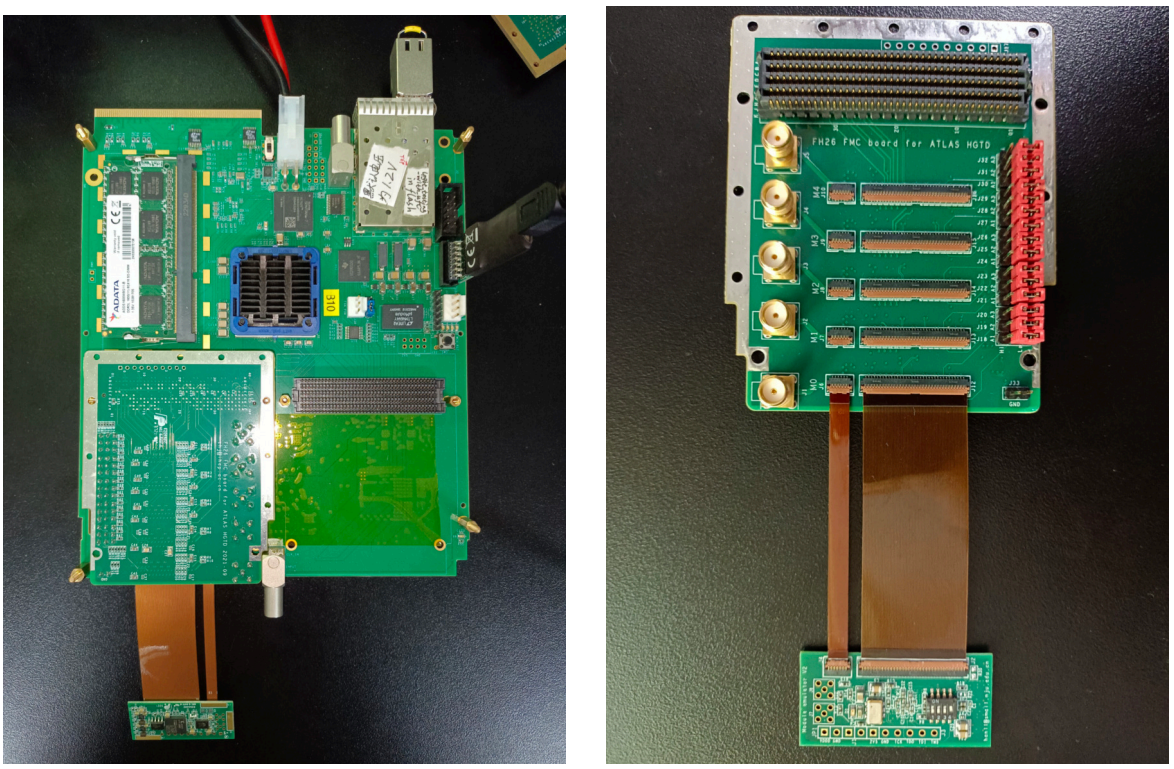


Figure B.3: The setup for the FLASH memory and FPGA basic functions test of module emulator.

daughter board, a dedicated experiment is set up to test all of them. The strategy of this test setup is to establish as much data transmission paths as possible that start from the module emulator FPGA and travel all the way to the final DAQ board, during which they successively go through the FPGA pins, 71-pin FH26W connector on module emulator, then 71-pin FH26W connector and FMC connector on FH26W daughter board. Figure B.4 shows two pictures of this setup.

Figure B.4(b) shows the top view of the combination of one module emulator and one FH26W daughter board. The module emulator is attached to the FH26W daughter board through two commercial flat cables, which are the substitutes of the FLEX cables that are still in design stage at the time of joint test. The left cable is supposed to transfer high voltages, but they will be not be applied in this test and the following joint test. The right cable holds all the transmission lines that bridge the two boards together both electrically and mechanically. The first one of the five slots is occupied in this picture, but other slots can also be taken one by one, in such a way, all the five FH26W connectors can be individually tested at the end. As can be seen, there are two variables in the test, one is about the forty module emulators under the test and the other one is about the five 71-pin FH26W connectors. Therefore, we should keep one variable unchanged while going through another variable. The adopted procedure is like this, the first slot of a certain FH26W daughter board remains unchanged, and all the forty module emulators are successively connected to it and get tested. When all the module emulators are done, a module emulator with good performance is chosen and remains unchanged. It will sequentially go through all the other FH26W connector slots not only one the current FH26W daughter board but also on other FH26W daughter boards that require connectivity test. Concerning the power supply of the module emulators, since the FH26W daughter board can only get the 1.2 V power supply from the FMC dedicated power pins, which are only routed to the first and second FH26W



(a) Global view of the test setup

(b) Top view of the module emulator and FH26W daughter board combination

Figure B.4: Pictures of the setup for the connectivity test of both the module emulator and FH26W daughter board.

connector slots, the module emulators that are connected to the other three slots have to be supplied through their external power input connector with the use of the custom-made interface clip.

Figure B.4(a) shows the a global view of the setup. As can be seen, again the uFC is used as the DAQ board of the test system. On the lower side, the combination of the module emulator and FH26W daughter board is inserted into the left FMC connector of uFC. As usual, the uFC is supplied by a 12 V voltage through the write power connector on its top side and a Xilinx JTAG programmer is inserted into its JTAG connector for FPGA programming and configuration. In order to initiate signals from module emulator side and receive them by the uFC, dedicated firmwares need to be developed and implemented into the FPGAs of module emulator and uFC. The development of the firmwares is actually based on the firmware for module emulator and its DAQ counterpart as described in Section 5.4.1, which can be just verified in virtue of this experiment. Figure B.5 shows the block diagram of the firmwares and connectivity of the setup.

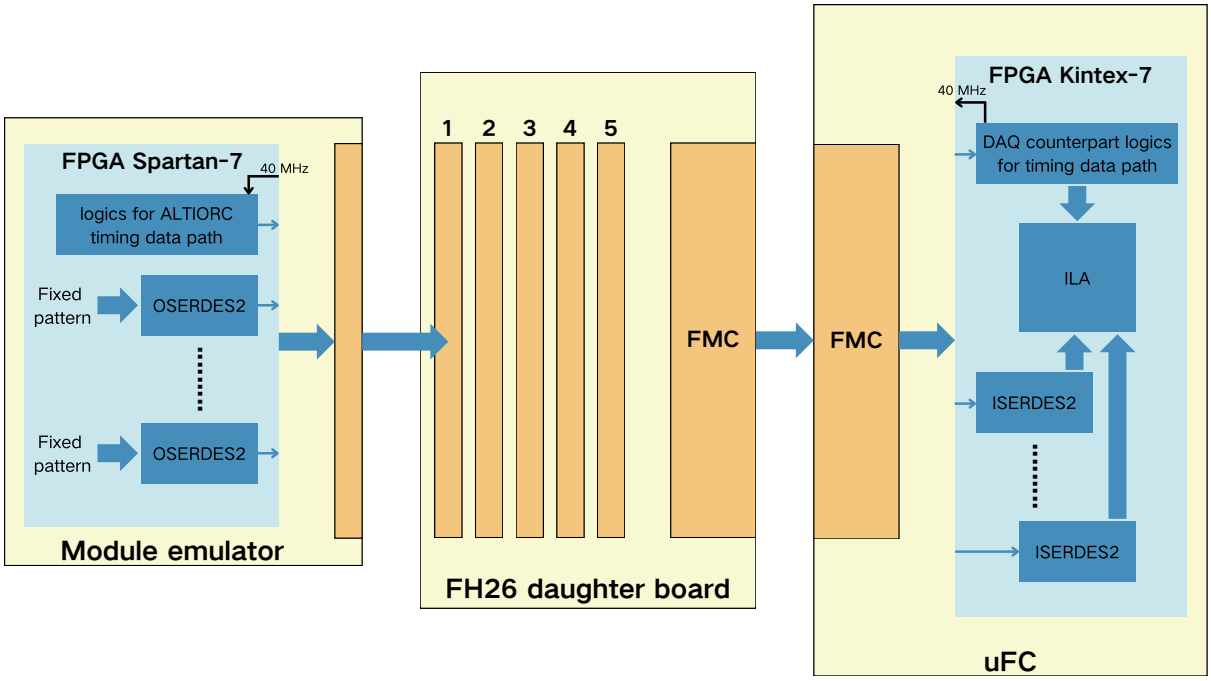


Figure B.5: The block diagram of the firmwares and connectivity of the setup.

For the firmware implemented in the module emulator FPGA, it aims to generate sufficient data streams to feed into the FPGA pins that are assigned for key signals such as the timing data, luminosity data, fast command, slow control, clock and monitoring signals. Among the data streams, one of them is generated by the logics for ALTIROC timing data path, then serialized and sent to the dedicated FPGA pins while the remaining ones are from fixed patterns, which are further serialized by the OSERDES2 logic and sent to related FPGA pins. After the long journey of passing through several connectors, those serialized data streams arrive at the corresponding pins of the FPGA on uFC board. The serialized

timing data stream is handled by the corresponding DAQ counterpart logic and reverted to the original timing data that is being monitored by the ILA logic. The other serialized data streams are deserialized by the ISERDES2 logic into the original fixed patterns, which also go to ILA logic. It's worth mentioning that, in order to synchronize the clocks that run on module emulator FPGA and uFC FPGA, the clocks on module emulator side are originated from a $40\ MHz$ clock from the uFC side. Based on the correctness of the recovered original data displayed on the host computer, the connectivity of those FPGA pins and also connector pins can be deduced.

Concerning the results of the connectivity test, firstly, it's found that there is no connectivity issue for all the connectors of the twenty FH26W daughter boards that go through the test, but for the module emulators, although there is also no connectivity issue for the 71-pin FH26W connector of each tested module emulators, it's found that nine out of forty module emulators suffer from insufficient soldering for a few pads of their FPGAs. Those nine problematic module emulators, together with the previous two boards that have FLASH memory issues, are sent back to the manufacturer for repair. Fortunately, all the issues are solved and those module emulators are sent to us.

Publications and Conference Talks

Publications with significant contributions:

- L. Han, L. Zhang, J. Zhang, S. Chen, M. Qi, Z. Liang, *Demonstration system of the HGTD peripheral electronics boards for ATLAS phase II upgrade*, NIM A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 1045, 2023, 167651, ISSN 0168-9002, https://doi.org/10.1016/j.nima.2022.167651.
- L. Han, L. Zhang, J. Zhang, S. Chen, M. Qi, Z. Liang, *The isolated USB programmer board for lpGBT configuration in ATLAS-HGTD upgrade*, 2022 JINST 17 C03030.
- ATLAS Collaboration, *Search for the non-resonant production of Higgs boson pairs via gluon fusion and vector-boson fusion in the $b\bar{b}\tau^+\tau^-$ final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, ATLAS-CONF-2023-071, 2023.

Conference posters and proceedings:

- **International conference:** *Isolated USB Programmer for LpGBT (UPL) for the ATLAS-HGTD upgrade*, TWEPP 2021 Topical Workshop on Electronics for Particle Physics, on-line, Sep 2021.
- **International conference:** *Demonstration System of the HGTD Peripheral Electronics Board (PEB) for ATLAS Phase II Upgrade*, 15th Pisa Meeting on Advanced Detectors, La Biodola - Isola d'Elba (Italy), May 2022.
- **National conference (Award for best poster and flash talk):** *Emulator System of the HGTD Peripheral Electronics Board (PEB) for ATLAS Phase II Upgrade*, The 7th China LHC Physics Workshop (CLHCP 2021), Nanjing, China, Nov 2021.
- **National conference:** *Legacy search for the non-resonant production of Higgs boson pairs via gluon fusion and vector-boson fusion in the $b\bar{b}\tau^+\tau^-$ final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, The 9th China LHC Physics Workshop (CLHCP 2023), Shanghai, China, Nov 2023.

Acknowledgement

致谢

完成这篇博士论文，我要感谢许多人的支持和帮助。在此，我谨向他们致以最诚挚的谢意。首先，我要衷心感谢我的导师陈申见教授。感谢您在我整个博士研究过程中的悉心指导和耐心支持。陈老师总是把学生放在心上，经常关心我的生活和规划。陈老师儒雅和蔼，循循善诱，常感受到轻松愉悦的科研氛围，能够在陈老师的指导下进行科研是我的幸运。

还要感谢我的导师张雷教授。张老师对科研总是一丝不苟、充满热情，这也一直激励着我对自己严格要求。经常看到张老师能够高效地解决科研上的复杂问题，这逐渐培养了我的科研和解决问题的能力。我的博士论文的完成离不开张老师对我科研上细致的指导和关心。感谢高能所的张杰老师，引领我入门硬件的研究，并且指导我在 HGTD 电子学的研发中做出成果。张杰老师对论文的硬件部分提出了宝贵的建议，这使得论文的质量更上一层楼。

感谢晓忠师兄对我在入门 ATLAS 实验时进行的指导和帮助，感谢 Antonio 博士后经常解答我在物理分析中碰到的疑惑。感谢博文、瀚非和汇润，我们一起讨论和解决科研以及生活上遇到的各种问题，度过了许多难忘的时光。感谢昱昊同学，我们一起去北京以及 CERN 做实验，互相帮助、解决遇到的问题。感谢久野良孝教授邀请我去大阪大学交流学习，感谢吴琛师兄在此期间带我了解 COMET 实验并指导我做了相关研究。感谢读博期间其他的同门和朋友，你们的无私帮助，我将铭记在心。

最后，我要感谢我的爷爷奶奶、爸爸妈妈和弟弟，感谢你们对我的无私支持，你们无条件的爱与关怀使我始终坚定前行。