Tomáz Antonio Bortoletto Giansante

A Study of Invisible Higgs Decays at a Future e^+e^- Collider Using the $Z(\mu^+\mu^-)$ Recoil Spectrum $\sqrt{s} = 250\,\text{GeV}$

Brazil

3 of March, 2026

Tomáz Antonio Bortoletto Giansante

A Study of Invisible Higgs Decays at a Future e^+e^- Collider Using the $Z(\mu^+\mu^-)$ Recoil Spectrum $\sqrt{s} = 250\,\text{GeV}$

Master Thesis for the Brazilian Center of Research in Physics, Particles Physics.

Brazilian Center For Research In Physics – CBPF

Ministério da Ciência, Tecnologia e Inovação – MCTI

Graduate Program

Supervisor: Carsten Hensel

Brazil

3 of March, 2026

Tomáz Antonio Bortoletto Giansante

A Study of Invisible Higgs Decays at a Future e^+e^- Collider Using the $Z(\mu^+\mu^-)$ Recoil Spectrum $\sqrt{s} = 250\,\text{GeV}$ / Tomáz Antonio Bortoletto Giansante. – Brazil, 3 of March, 2026-

134p. : il. (algumas color.) ; 30 cm.

Supervisor: Carsten Hensel

Dissertation (Master) – Brazilian Center For Research In Physics – CBPF
Ministério da Ciência, Tecnologia e Inovação – MCTI
Graduate Program, 3 of March, 2026.

1. Invisible Higgs Decay. 2. International Linear Collider. 2. Key4hep. I. Carsten Hensel. II. Brazilian Center for Research in Physics. III. Ministério da Ciência, Tecnologia e Inovação. IV. Master's Degree

"A STUDY OF INVISIBLE HIGGS DECAYS AT A FUTURE e^+e^- COLLIDER USING THE Zmu+mu-RECOIL SPECTRUM sqrt(s)=250 GeV"

TOMÁZ ANTONIO BORTOLETTO GIANSANTE

Dissertação de Mestrado em Física apresentada no Centro Brasileiro de Pesquisas Físicas do Ministério da Ciência Tecnologia e Inovação. Fazendo parte da banca examinadora os seguintes professores:

Carsten Hensel - Orientador/CBPF

Helena Brandão Malbouisson – UERJ

Dilson de Jesus Damião – UERJ

Rio de Janeiro, 14 de abril de 2026.

This work is dedicated to everyone
that helped me become a Physicist.

Acknowledgements

Firstly, I wish to express my sincere appreciation to the Brazilian Center for Research in Physics (CBPF) for accepting me as a master's student and providing an excellent research environment.

I would like to my love and appreciation for my closest friends-Mariana Moraes, Felipe Mesquita, Ivan Gentil, Guilherme Baccari, Matheus Alves, Luiz Henrique and many others. I owe a special debt of gratitude to my friend Rodrigo Manhães, whom I shared my stay at Rio de Janeiro, where we learned and grew together with unforgettable experiences.

Special thanks to my father, Antonio Eduardo, my brother, Miguel Giansante and my mother, Cláudia Bortoletto for helping and motivating me during my entire Physics journey, from my undergraduate degree at Universidade de São Paulo to my master's degree at CBPF. Supporting me in my most difficult moments leading me to finish my degree.

I also want to thanks Deutsches Elektronen-Synchrotron (DESY), for providing me access to their ecosystem for the Key4hep simulations that made this thesis possible.

My utmost thanks for Carsten Hensel for being my supervisor, whose mentorship and insightful feedback shaped the entirety of this work.

I am deeply thankful to CNPq for awarding me an scholarship.

"Investigation may be likened to the long months of pregnancy, and solving a problem to the day of birth. To investigate a problem is, indeed, to solve it."

- Mao Tsetung

Abstrato

A busca pelo Decaimento Invisível do Higgs (IHD) constitui uma sonda poderosa e independente de modelo para investigar física além do Modelo Padrão (SM), com particular relevância para cenários de matéria escura nos quais o bóson de Higgs se acopla a partículas fracamente interagentes.

Nesta tese, o objetivo é demonstrar a viabilidade da estrutura Key4hep usando o Decaimento Invisível do Higgs como um benchmark de fluxo de trabalho. O fluxo de trabalho inovador Key4hep desenvolvido pelo autor resultou em uma cadeia de análise simulada que permite estimar a sensibilidade ao canal de decaimento invisível do Higgs $h \to \text{inv}$ é avaliada no canal de *Higgsstrahlung* $e^- e^+ \to Zh$, a uma energia no centro de massa de $\sqrt{s} = 250\,\text{GeV}$, assumindo uma luminosidade integrada de $\mathcal{L} = 250\ \text{fb}^{-1}$ e feixes polarizados correspondentes ao estágio inicial de operação do International Linear Collider (ILC) (1, 2). É realizada uma simulação completa de Monte Carlo dos processos de sinal e dos backgrounds do SM, com foco no modo de decaimento leptônico $Z \to \mu^+\mu^-$ e no IHD $h \to \text{inv}$ com a suposição final do MP de $\text{inv} = 2\nu + 2\bar{\nu}$.

A análise emprega a técnica de massa de recuo independente de modelos para identificar decaimentos invisíveis do Higgs, independentemente de seus produtos de decaimento. Uma seleção de eventos baseada em cortes é desenvolvida, inspirada em buscas anteriores por decaimentos invisíveis do Higgs tanto no LHC (3) quanto no ILC (2, 4, 5, 6). Os cortes são definidos com base em variáveis cinemáticas, como massa de recuo, energia ausente e momento transversal aplicados para maximizar a sensibilidade do sinal enquanto suprimem os backgrounds.

Sob essas hipóteses, é realizada uma projeção simplificada de sensibilidade por meio de uma abordagem baseada em contagem de eventos. A partir dos rendimentos selecionados de sinal e background, estima-se uma sensibilidade esperada aproximada aos decaimentos invisíveis do Higgs, correspondente a um alcance na razão de ramificação de

$$\text{BR}(h \to \text{inv}) < 4.5\% \tag{1}$$

Essa estimativa é derivada de uma análise baseada em cortes, com tratamento aproximado das incertezas, e não constitui um limite formal obtido por meio de um ajuste de perfil de verossimilhança (*profile-likelihood*) ou de um procedimento de exclusão CL_S. Esses resultados demonstram o forte potencial do ILC para observar ou impor restrições significativas aos decaimentos invisíveis do Higgs, fornecendo testes rigorosos de cenários de nova física além do SM.

Palavras-chave: Particle Physics. Key4hep. Linear Collider. Higgs. Invisible Decay. Neutrinos. Dark Matter. Física de Partículas. Key4hep. Aceleradores Lineares. Higgs. Decaimento Invisível. Neutrinos. Matéria Escura.

Abstract

The search for Invisible Higgs Decay (IHD) provides a powerful and model-independent probe of physics beyond the Standard Model (SM), with particular relevance to dark matter scenarios in which the Higgs boson couples to weakly interacting particles.

In this thesis, the goal is to demonstrate the viability of the Key4hep framework using the Invisible Higgs Decay as a workflow benchmark. The innovative Key4hep workflow developed by the author, resulted in a simulated analysis chain that allows the estimation of the sensitivity to the invisible Higgs decay channel $h \to \text{inv}$ evaluated in the *Higgstrahlung* channel $e^- e^+ \to Zh$ at a Center-of-Mass energy of $\sqrt{s} = 250\,\text{GeV}$, assuming an integrated luminosity of $\mathcal{L} = 250\ \text{fb}^{-1}$ and polarized beams corresponding to the initial operational stage of the International Linear Collider (ILC) (1, 2). A full Monte Carlo simulation of signal and Standard Model background processes is performed, focusing on the leptonic decay mode of the $Z \to \mu^+\mu^-$ and the IHD $h \to \text{inv}$ with the final state SM assumption of $\text{inv} = 2\nu + 2\bar{\nu}$.

The analysis employs the model-independent recoil mass technique to identify invisible Higgs decays regardless of the Higgs decay products. A cut-based event selection is developed inspired by previous invisible Higgs decays searches at either the LHC (3) or the ILC (2, 4, 5, 6). The cuts are based on kinematic variables such as recoil mass, missing energy and transverse momentum are applied to maximize signal sensitivity while suppressing backgrounds.

Under these assumptions, a simplified sensitivity projection is performed using a counting-based approach. From the selected signal and background yield, an approximate expected sensitivity to invisible Higgs decays is estimated, corresponding to a branching ratio reach of

$$\text{BR}(h \to \text{inv}) < 4.5\% \tag{2}$$

This estimate is derived from a cut-based analysis with an approximate treatment of uncertainties and does not a full profile-likelihood limit or CL_S exclusion limit. These findings demonstrate the strong potential of the ILC to either observe or significantly constrain invisible Higgs decays, thereby providing stringent tests of new physics scenarios beyond the SM.

Keywords: Particle Physics. Key4hep. Linear Collider. Higgs. Invisible Decay. Neutrinos. Dark Matter.

List of Figures

List of Tables

List of abbreviations and acronyms

BSM	Beyond the Standard Model
CBPF	Brasilian Center for Physics Research
CERN	European Organization for Nuclear Research
DESY	Deutsches Elektronen-Synchrotron
ECAL	Electromagnetic calorimetry
ELOG	ELOG gendata
FCC	Future Circular Collider
FoM	Figure of Merit
ggF	Gluon-Gluon Fusion
GUT	Grand Unified Theory
HB	Higgs Boson
HCAL	Hadronic Calorimeter
HEP	High Energy Physics
HF	Higgs Field
ILD	International Large Detector
ILC	International Linear Collider
IHD	Invisible Higgs Decay
ISR	Initial State Radiation
LHC	Large Hadron Collider
MC	Monte Carlo
MVA	Multivariate Analysis
PDG	Particle Data Group
QFT	Quantum Field Theory

ROOT	Rapid Object-Oriented Technology
SiD	Silicon Detector
SM	Standard Model
SSB	Spontaneous Symmetry Breaking
TM	Transverse Momentum
PS	Production Summary
PFA	Particle Flow Algorithm
VBF	Vector Boson Fusion
VEV	Vacuum Expectation Value
YC	Yukawa Coupling

List of symbols

B	Background yield
b	Bottom quark
c	Speed of light in vacuum
d	Down quark
E	Energy
E_{miss}	Missing energy
e^-	Electron
e^+	Positron
g	Gluon
H	Higgs boson
$m_{\mu\mu}$	Di-muon Invariant mass
m_{recoil}	Recoil mass
n	Neutron
N_{gen}	Number of generated events
N_{sel}	Number of selected events
p	Proton
p_T	Transverse momentum
R	Background rejection
S	Signal yield
t	Top quark
W	W boson
w	Event weight
Z	Z boson

ϵ	Signal efficiency
γ	Photon
Γ	Decay Width
δ	Uncertainty
ℓ	Charged lepton
$\mathcal{L}$	Integrated luminosity
μ	Muon
ν_e	Electron neutrino
ν_μ	Muon neutrino
ν_τ	Tau neutrino
σ	Cross section
$\sqrt{s}$	Center-of-mass energy
τ	Tau lepton

Contents

1 Introduction

The discovery of the Higgs boson at the CERN Large Hadron Collider (16) confirmed the last missing piece of the Standard Model (SM). Theory shows that the Higgs boson has an essential role in the evolution of the Universe from it's very beginning to the physics we see in our experiments today (17), but despite its remarkable success, the SM does not explain several fundamental observations specially at very high energy scales, including the nature of dark matter, neutrino oscillations, particles that weakly couples to SM particles, or particles with a compressed mass spectrum and etc. One of the most sensitive and model-independent probes of physics Beyond the Standard Model (BSM) is the search for invisible decays of the Higgs boson. The SM predicts that the Higgs rarely decays invisibly, an expected branching ratio (BR) of 0.1%, with a final state consisting of two pairs of neutrinos $h \to 2\nu + 2\bar{\nu}$, any deviation from the tiny SM prediction would be a clear indication of new physics, such as Higgs-portal dark matter or other weakly interacting hidden particles (2, 4, 5, 1).

The Higgs sector is where it's expected the majority of new physics beyond the SM, improvement in the precision levels of Higgs experiments requires upgrades and construction of current and future particle accelerators (1). A future electron–positron collider operating as a Higgs factory provides an ideal environment for this search. The International Linear Collider (ILC) (18), planned to operate initially at a center-of-mass energy of $\sqrt{s} = 250\,\text{GeV}$, maximizes the production of Higgs bosons via the *Higgsstrahlung* process $e^+e^- \to Zh$, considering also that the momentum resolution degrades linearly with momentum. In contrast to hadron colliders, the clean environment of e^+e^- collisions combined with the improved beam polarization that increases the *Higgsstrahlung* cross-section for $e^+_L e^-_R$ (4). The combination of all these facts allows precise and model-independent measurements of Higgs properties.

A particularly powerful technique at $\sqrt{s} = 250\,\text{GeV}$ is the recoil mass method. By reconstructing only the Z boson—especially in the clean leptonic channel $Z \to \mu^+\mu^-$, the presence of the Higgs boson can be inferred from conservation of four-momentum, independently of how the Higgs decays. This makes the recoil method uniquely suited for probing invisible Higgs decays (IHD), since the analysis does not rely on detecting the Higgs decay products directly.

In this thesis, the sensitivity to the IHD channel is evaluated for the ILC operating at $\sqrt{s} = 250\,\text{GeV}$ using the $Z \to \mu^+\mu^-$ recoil channel, the $\mu^+\mu^-h$ channel is advantageous as it has a sharper signal peak, hence better mass resolution the e^+e^-h channel that suffers from *bremsstrahlung* (4) combined with the fact that larger center-of-mass energies

the *bremsstrahlung* affects the tail of recoil mass for signal and background processes, therefore making the $\sqrt{s} = 250\,\text{GeV}$ ideal.

A full simulation and analysis chain is implemented within the Key4hep framework (19), including event reconstruction, preselection, sequential cut optimization, and statistical evaluation. The IHD is treated as signal, while all SM processes producing muon pairs at this energy are considered background. The final sensitivity is quantified using a Figure of Merit (FoM) and signal-to-background ratio S/B after a complete cutflow workflow.

To study the future of IHD experiments, the Key4hep framework was used to simulate the collisions, that simulates HEP through Monte Carlo Simulation. The Key4hep work-flow is divided into a few crucial pieces: the input files, either `ROOT` or `slcio`, that contain the information of the desired event, a Gaudi algorithm, a data processing framework, that allows the selection of the samples produced for further analysis and write it into a `ROOT` file, a `python` steering file, that serves as the Key4hep interface and it connects the input files and the Gaudi algorithms. The simulated events selection was only done if a pair of muons $\mu^+\mu^-$ was found, considering the IHD as the Signal and other events that occur at 250 GeV as background.

Background samples have characteristics similar or equal topology as the signal, all events that produce at least a muon pair of opposite charge $\mu^+\mu^-$ will enter into the simulated samples. The cutflow analysis paradigm is used across HEP as a selection criteria, with the purpose of removing the background without significant loss of signal samples. Using the established SM knowledge, the simulated events are cut based on thresholds based on the expected behavior of the IHD, the cuts were based on the following variables: presence of the di-muon $\mu^+\mu^-$, Multiplicity, the amount of charged particles, the Transverse Momentum (TM) of each muon p_T, the muon pair energy $E_{\mu\mu}$, the invariant mass, to be equal to the Z boson mass m_Z, the recoil mass to be equal to the Higgs boson mass m_h, the dimuon TM p_T and missing energy, E_{miss}. Later, the final simulated events are analyzed using statistical tools, the Figure of Merit (FoM) and a a simplified projected sensitivity estimate as proof-of-principle for future analysis.

1.1 Structure of the Thesis

The present work is organized into six chapters.

- Chapter 1 - Introduction. Overview of the contemporary development of the Higgs mechanism within SM and it's contextualization inside the current development of the Future Circular Collider and specially ILC in the current particle-physics research.
- Chapter 2 - The Higgs Boson. Reviews the theoretical development of the Higgs field and its associated particle, the Higgs particle, chronicles of the first experimental observations of the Higgs particles at the Large Hadron Collider. The chapter also delineates Higgs role and decay channels within the SM. Lastly, the limitations of the SM in describing Higgs-related phenomena.
- Chapter 3 – The International Linear Collider. Begins with a historical survey of accelerator technologies, emphasizing the conceptual distinctions between linear and circular machines. It then describes the principal components of the ILC, the International Large Detector and the Silicon Detector, and analyses how these facilities advance the study of the IHD in comparison to current experimental programs like the Large Hadron Collider.
- Chapter 4 – Simulation and Data Analysis. Details the Monte-Carlo simulation framework within Key4Hep employed to generate synthetic ILC samples, overview of background files provided by Deutsches Elektronen-Synchrotron (DESY) explains the architecture of the analysis code, and presents the methodology used to extract physical observables. is the analysis of the samples generated through the simulation of the experiment, the code structure and purpose.
- Chapter 5 – Results. Summarizes the principal findings obtained from the simulated simulated samples, including quantitative assessments of the expected IHD and simulated detector performance metrics.
- Chapter 6 – Conclusions and Outlook. Finishing thoughts, reflections on the implications of the results in regards current and future Higgs-boson invisible decays studies. Subsequent experimental and theoretical investigations for SM and BSM physics.

2 The Higgs Boson

In the earliest moments of the Universe, after the Big Band and even before the first second had passed, particles were devoid of mass and all forces were unified, according to modern theory. This meant that all particles sped around the speed of light ($c = 299\,792\,458 \mathrm{m\,s^{-1}}$). Only when the Universe expanded and cooled, until it reached a state that almost all particles instantly acquired mass through an field, the Higgs field, a field which permeates the entirety of space, associated with its own particle, the Higgs boson h, that through interaction with other particles, gives birth to the diverse masses of particles observed today (20). Currently, physics search for reason for why particles have acquired the specific values for their masses as they are observed in Nature (17). In spite of this open question, it's due to the Higgs mechanism that the formation of stars, planets, and, ultimately, life as we know it became possible.

The Higgs Boson in conjunction with the Higgs mechanism was proposed in 1960s by François Englert, Robert Brout, Peter W. Higgs, Gerald Guralnik, C. R. Hagen and Tom Kibble but the particle and field name ended up being attributed to Peter Higgs. After decades of development in particles physics experiments, the Higgs boson (h) was eventually discovered in experiments conducted by ATLAS and CMS collaborations at CERN (European Organization for Nuclear Research) in 2012 (20), rewarding François Englert and Peter W. Higgs the Nobel Prize in physics in 2013 (21).

According to Quantum Field Theory (QFT), the Higgs boson is actually a excitation of the Higgs Field (HF), a wave in a quantum field. Unlike any other SM particles, the Higgs is one-of-a-kind particles, it has zero spin, no electric charge and no color charge. To understand the Particle, first a brief synthesis of the Field is required.

2.1 Higgs Field And Boson

Particles interact with the HF, acquiring mass as a consequence (20), however due to the high energy density of the early universe the field was constantly excited. The excited field didn't give mass to the particles until reaching a certain point of expansion of the universe, where the field spontaneously settled into a particular value in the lowest energy state, this mechanism is the Higgs-Brout-Englert-Guralnik-Hagen-Kibble mechanism, also named as the Spontaneous Symmetry Breaking (SSB) or the Higgs mechanism (9). The matter-antimatter asymmetry in the Universe involves a possible first order phase transition with the Higgs Vacuum Expectation Value (VEV) generated in the early Universe. Bubbles with Higgs condensate would be created and expand at the speed

of light (17).

2.1.1 The Standard Model

Developed in the early 1970s, the Standard Model (SM) has successfully explained almost all experimental results and precisely predicted a wide variety of phenomena. Despite the overwhelming success of the SM, gaps within the model are the focus of current and future physics experiments. Firstly, the SM can't describe things like: Dark Matter, behavior at the Grand Unified Theory (GUT) scale, the existence of neutrinos masses and the V-A structure of the SM (as the V-A structure requires neutrinos to be massless). Secondly, some SM predictions have not yet have been confirmed experimentally like the IHD. The latter type is the search for proof of the SM predictions, as the IHD is predicted by the SM, measuring it requires improved precision to expose if any violation of the SM prediction is wrong. Figure 1 has all seventeen fundamental particles predicted by the SM, fundamental means that the particle it's made only of itself, the anti-matter variants of these particles are also expected, any other particle not represent in the figure is not a fundamental particle. Particles in the SM model are divided into two categories, fermions, matter particles and bosons, force carriers.

Bosons are force carriers: the gluon is the strong force carriers, the photon, the electromagnetic force carrier and the Z and W bosons, the weak force carriers, the gravity force carrier is absent from the SM, some BSM physics predict the gravity force carrier to be the graviton but there's still to be confirmed. There two types of bosons, vector bosons, where their fields interactions are vectorial, with a direction and magnitude and scalar bosons, with only the Higgs boson present, where its fields interactions don't have direction but only magnitude, therefore, a scalar value.

2.1.2 The Higgs mechanism

Assuming the QFT convention that the speed of light is $c = 1$ and the plank constant is $\hbar = 1$. The Higgs doublet is represented as:

$$\phi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix} \tag{2.1}$$

To achieve the necessary symmetry breaking, the Higgs mechanism employs scalar fields, the Higgs bosons. The Lagrangian that describes these scalars is defined by the equation:

$$\mathcal{L}_s = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi) \tag{2.2}$$

where D_μ is the covariant derivative. The potential term $V(\phi^\dagger \phi)$ is gauge invariant and is represented as:

$$V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{2.3}$$

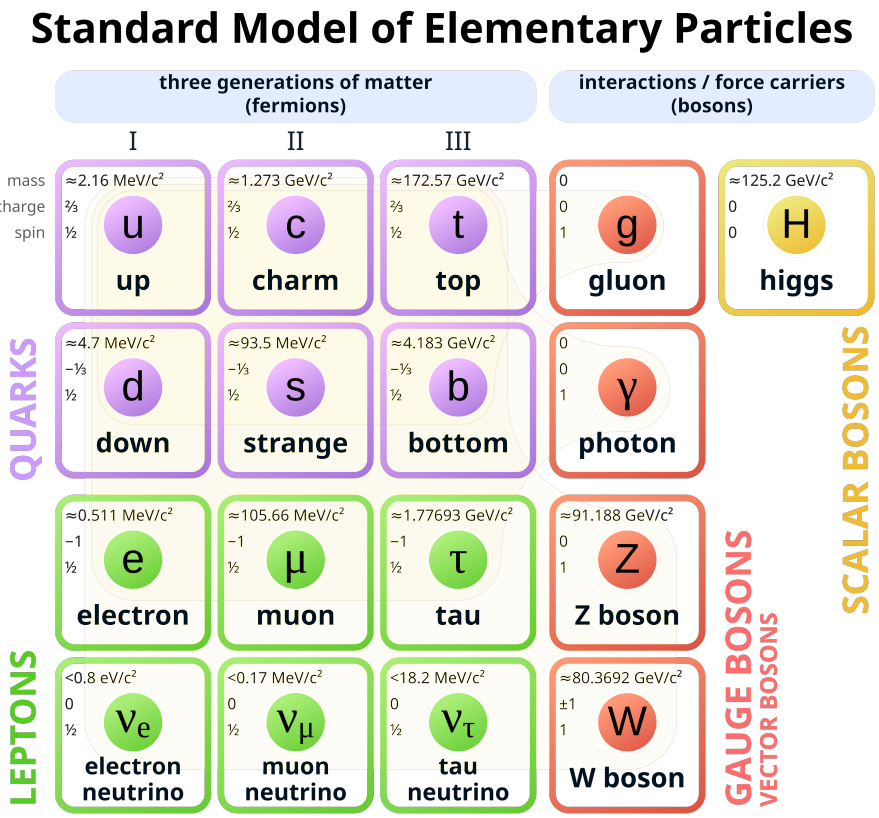


Figure 1 – Elementary particles of the Standard Model (7).

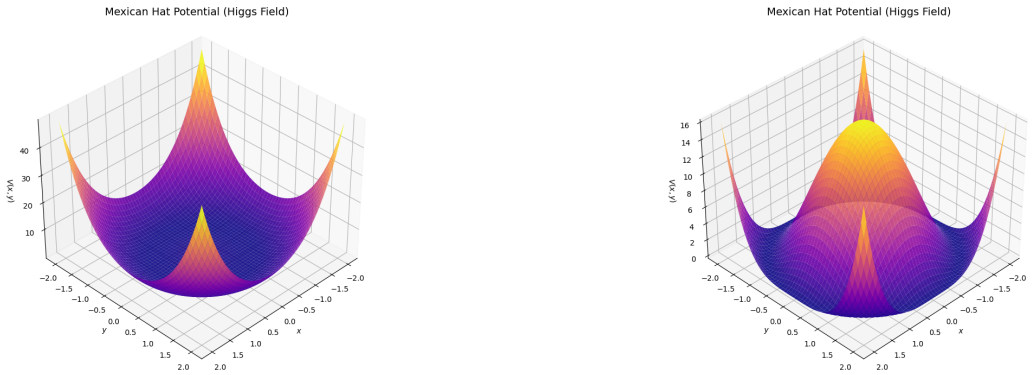
the parameters μ^2 and λ are both real constants, λ needs to be positive to ensure a stable vacuum. The theory requires renormalization, which means that higher powers of $\phi^\dagger\phi$ are not allowed (20). The SSB occurs when the scalar doublet ϕ develops a VEV:

$$\phi_0 = \langle 0|\phi|0\rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \tag{2.4}$$

where $v = \sqrt{-\mu^2/\lambda}$. The Higgs field gives mass to the Z and W bosons (Higgs mechanism), the quarks and charged leptons (via Yukawa couplings) but not all elementary particles through the Higgs mechanism, the gluon and the photon are massless, Table 1 has all particles that couple with the Higgs Field. The SM doesn't predict why neutrinos have mass, whereas, the experimental proof of neutrino oscillations indicates that they do have a very small mass, however, it cannot be assumed that the neutrino mass comes from the Higgs mechanism.

2.2 Production

There are four main production mechanisms of the Higgs boson, each one is dependent on the initial state particles that are collided. The LHC does proton-proton



(a) Before Symmetry Breaking.

(b) After Symmetry Breaking.

Figure 2 – Higgs potential $V(|\phi|)$ in field space (3D) for two choices of the parameter μ^2. Figure 2a is for $\mu^2 > 0$, the minimum occurs at $v = 0$ and the symmetry is unbroken. Figure 2b is for $\mu^2 < 0$, the potential develops a degenerate set of minima at $v = \sqrt{-\mu^2/\lambda}$, leading to spontaneous symmetry breaking. Source: Author's own creation.

Table 1 – Particle Masses and their coupling to the Higgs Field according to the SM.

Category	**Particles**	**Mass from Higgs Field**
Particles that acquire mass from the Higgs field		
Quarks	u, d, c, s, t, b	Yes
Charged Leptons	e^-, μ^-, τ^-	Yes
Weak Gauge Bosons	W^+, W^-, Z^0	Yes
Particles that do NOT get mass from the Higgs field		
Photons	γ	No (massless)
Gluons	g (8 types)	No (massless)
Neutrinos	ν_e, ν_μ, ν_τ	Uncertain

hadron collisions, while the FCC and ILC will collide electron-positrons, resulting in a difference of Higgs production mechanisms for each experiments. The following sections giving a brief overview of most of them and going in depth for the *Higgstrahlung* process.

For hadron colliders like the LHC, the most dominant production mechanism is the gluon-gluon fusion via a top quark loop $(ggF)(gg \to h)$ (3), while the Vector Boson Fusion (VBF) process has its cross-section proportional to the increase in energy and takes over that of the *Higgstrahlung* process at $\sqrt{s} \gtrsim 400\,\text{GeV}$ (8), it has been previously studied at $\sqrt{s} = 7, 8, 13, 14$ TeV at the LHC.

For lepton colliders like the ILC, VBF, top quark pairs $(tth)(gg/q\bar{q} \to t\bar{t}h)$ and the *Higgstrahlung* are the major Higgs production modes. The *Higgstrahlung* (Zh) channel is the main Higgs produciton mode at $\sqrt{s} = 250\,\text{GeV}$, where it attains its maximum cross-section (8), this coincides with the first stage of operation of the ILC (1), therefore making the IHD a great benchmark to test the capabilities of the ILC.

Out of the four channels, the associated production channel Vh $(V = W, Z)$ is the

cleanest; Figures 3, 4, 5 and 6 are the Feynman diagrams of the all the Higgs productions modes are represented through their leading order. However, the Zh channel is more promising, as unlike Wh, the presence of two leptons in the final state, in this case, muons $\mu^+\mu^-$, from the Z boson decay (3) allows for better detection.

2.2.1 Gluon-Gluon Fusion

The Gloun-Gloun Fusion (ggF) is the dominant production channel at the LHC, responsible for 86% of the Higgs production at $\sqrt{s} = 8\,\text{TeV}$. Figure 3 is the Feynman diagram of the ggF, the middle triangle is the heavy-quark loop, in resume the final state of the decay is:

$$gg \to h \tag{2.5}$$

the complete decay chain for ggF is

$$gg \to t\bar{t} \to h \tag{2.6}$$

the Higgs produced also decays invisibly. However, at leading order, it's hard to detect because of soft missing TM ($\not{p}_T$), at higher orders of QCD, the Higgs can be produced in association with a single jet and has a considerably large missing TM along with a jet (3).

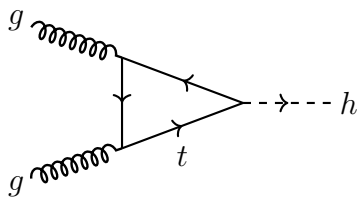


Figure 3 – Feynman diagram of Higgs boson production via gluon-gluon fusion ($gg \to t\bar{t} \to h$) mediated by a top-quark loop.

2.2.2 Vector Boson Fusion

In the VBF channel, the Higgs is produced throught vector boson fusion, where vector boson originate by radiating off two initial quarks along with two jets. A quark from each incoming direction radiates a heavy vector boson (V), which are either W or Z, the bosons the "fuse" to produce the Higgs boson. The initial quarks are slightly deflected and travel roughly along their initial directions, then they are detected as particles "jets" by the detector (3):

$$pp \to qqh \to 2 \text{ jets} + \not{p}_T \tag{2.7}$$

where the final state consists of the missing TM $\not{p}_T$ due to the presence of non-interacting particles and two jets in the forward and backward directions as seen in the VBF Feynman diagram at Figure 4.

The unique event topology results in a more effective removal of background, this makes VBF one of the most promising channels for the detection of an invisibly decaying Higgs (3).

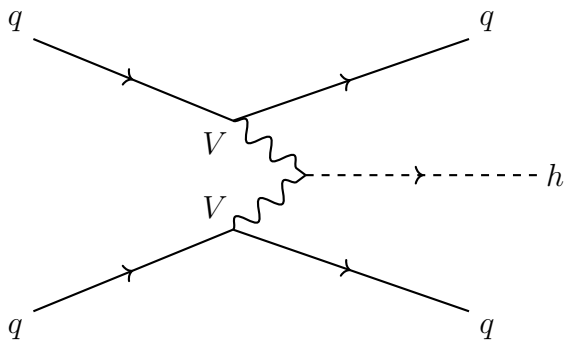


Figure 4 – Feynman diagram of Higgs boson production through vector boson fusion (VBF) in proton–proton collisions ($pp \to qqh$). In this process, two quarks emit virtual electroweak gauge bosons ($V = W^{\pm}, Z$), which fuse to produce the Higgs boson.

2.2.3 Associated Production

2.2.3.1 Top Quark Pair

The associated production of Higgs with a top quark pair is the process with the lowest cross-section out of the four production processes cited (3), where Figure 5 is the Feynman diagram of $t\bar{t}h$ and its decay looks like:

$$gg/q\bar{q} \to t\bar{t}h \tag{2.8}$$

The $t\bar{t}h$ channel has been studied in detail for $\sqrt{s} = 14\,\text{TeV}$ at the LHC (3).

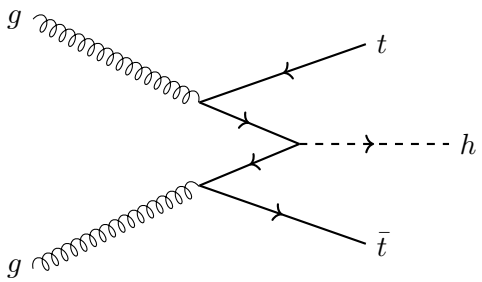


Figure 5 – Feynman diagram for Higgs boson production in association with a top–antitop quark pair ($e^+e^- \to t\bar{t}h$).

2.2.3.2 Z/W Boson

Particles accelerators like the FCC and ILC are planned to do collisions with e^-e^+ pairs, when an elementary fermion encounters its anti-fermion they can combine to form a virtual W or Z (20), such pair is capable of producing Higgs through consequent decays, $e^-e^+ \to h$. However the chances of such event is about eight times smaller than the production of Higgs in the $e^-e^+ \to Zh$ associative process (22).

The associated production, characterized by the presence of a Z and h boson at the same step, is the most common way to produce is through the electron-positron (e^+e^-) collision at $\sqrt{s} = 250\,\text{GeV}$ at which it achieves it's maximum cross-section and significant reduce in the *beamstrahlung* effect, the radiation of photons from electrons or positrons in one colliding bunch due to the field produced by the oncoming colliding bunch (1):

$$e^+e^- \to Z^* \to Zh \tag{2.9}$$

and in this channel, the Z boson can decay both leptonically, $Z \to \ell\bar{\ell}$, for the IHD, $\ell = \nu$, and hadronically, $Z \to b\bar{b}$ (3). Table 6 shows that the hadronic decay is the second muon-producing event with the highest cross-section σ.

The presence of an intermediate Z^* boson in equation 2.9, is necessary for the production of h, the superscript $*$ indicates that the Z is actually an virtual particle, also known as a off-shell particle, a virtual boson that caries sufficient energy with the potential of emitting a Higgs boson, a process know as *Higgstrahlung* (20), Figure 6 is the Feynman diagram, where it can be seen the intermediary virtual Z^*. Since the object of study is

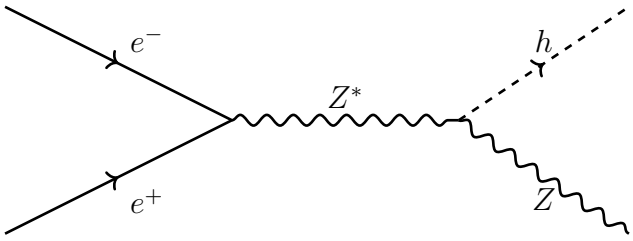


Figure 6 – Feynman diagram of the *Higgstrahlung* process ($e^+e^- \to Z^* \to Zh$), which is the dominant Higgs production mechanism at center-of-mass energies around 250 GeV.

Higgs but the Z boson is also present, it is said the Z boson is produced in association with the Higgs (22). A on-shell particle has to abide to the mass-shell condition:

$$E^2 = p^2 + m^2 \tag{2.10}$$

an *off-shell* particle doesn't follow this condition due to the fact that because they are intermediary and temporary particles, its lifespan is shorter (Δt) and therefore by the Heisenberg's Uncertainty Principle:

$$\Delta E \Delta t \sim \hbar \tag{2.11}$$

the uncertainty in the energy ΔE it's sufficiently large to allow the particle to reach energy levels higher than that was originally in the system, violating the mass-shell condition.

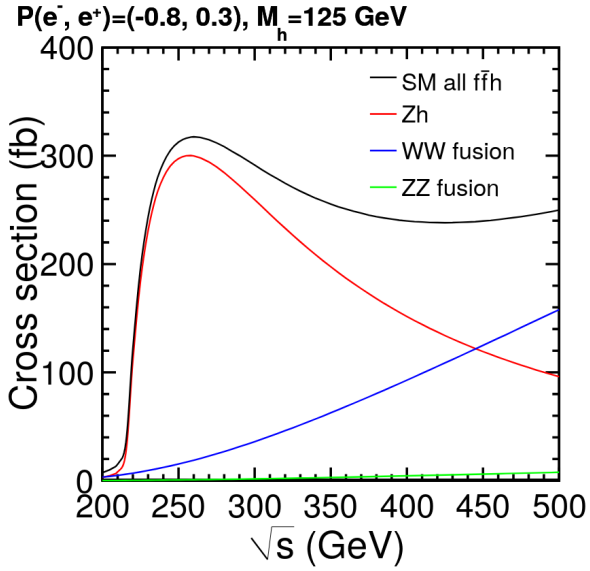


Figure 7 – Higgs production modes cross-sections, Zh *Higgsstrahlung* (red), vector boson fusion for WW (blue) and ZZ (green), sum of all production modes (black). (8)

2.3 Detection

Because h is hard to directly detect, the usual approach is to analyze its decay products. The decay width Γ of a particle is the inverse of it's lifetime and it can be divided into a sum of Γ_X, where Γ_X is the partial decay width for each specific decay mode $h \to X$, then a branching fraction is defined as the probability that a decaying particle will decay into a particular final state among all possible decay modes. The coupling of the Higgs boson to X can be written as:

$$\mathrm{BR}(h \to X) = \frac{\Gamma(h \to X)}{\Gamma_{\mathrm{total}}} \tag{2.12}$$

where Γ_{total} is the sum over all modes $\Gamma(h \to X)$ and it's equal to 1 (1). Experimentally, a BR measurement is similar to a cross-section measurement. Given a produced sample of N_{total} particles and a number N_i of final states corresponding to the decay chain i, the BR can be measured as:

$$\mathrm{BR}(h \to X) = \frac{N(h \to X)}{N_{\mathrm{total}}} \tag{2.13}$$

Table 2 has the SM BR predictions for various Higgs decay modes. Such decays are described by the Yukawa Coupling (23), it's possible to see that the highest branching

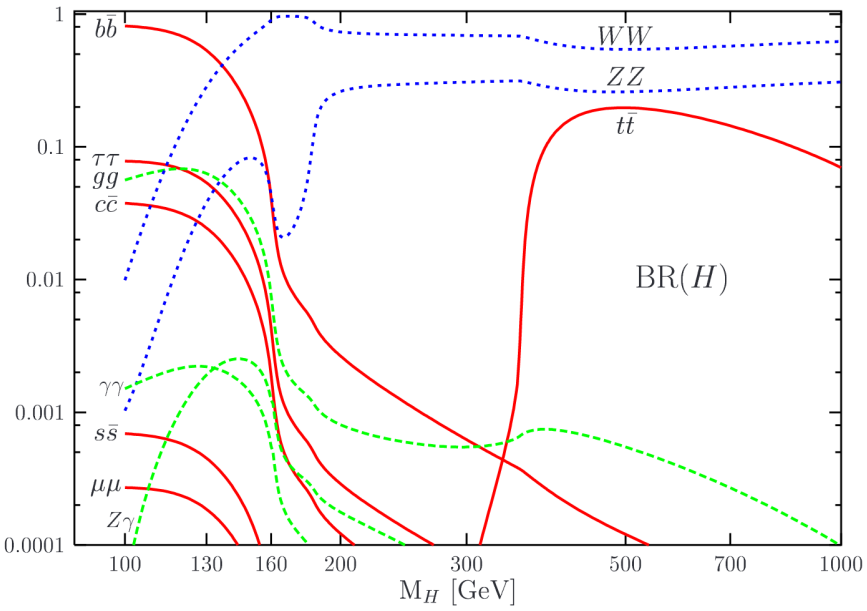


Figure 8 – The SM Higgs boson decay BRs as a function of M_h (9)

factors are with the most massive particles. The YC describes the interactions between the Higgs Field and the matter fields, the YC coupling between the Higgs boson the fermions, y_F, is proportional to the fermion masses m_F (17). The top quark, the heaviest quark, has the strongest coupling, whereas for the electron, being much lighter. The Higgs boson doesn't direct decay into neutrinos, as the SM doesn't contain coupling for that, justifying their absence from Table 2.

Table 2 – SM predictions for the Higgs decay branching fractions (12).

Decay Mode	BR (%)	Decay Mode	BR (%)
$h \to bb$	57.5 ± 1.9	$h \to ZZ^*$	2.67 ± 0.11
$h \to WW^*$	21.6 ± 0.9	$h \to \gamma\gamma$	0.228 ± 0.011
$h \to gg$	8.56 ± 0.86	$h \to Z\gamma$	0.155 ± 0.014
$h \to \tau\tau$	6.30 ± 0.36	$h \to \mu\mu$	0.022 ± 0.001
$h \to cc$	2.90 ± 0.35		

The Higgs Boson is a massive scalar boson, with mass of $m_h = 125\,\text{GeV}$, by scalar it means that the Higgs Field is a Scalar Field, where each point in the field is a number and has no direction (unlike the electromagnetic field that has direction) has no electric charge, no color charge and no spin, the only particle that has no spin in the SM. Therefore, it's a challenge to detect or at least directly, by knowing which particles it decays to, the Higgs decay chain can be reconstructed. Higgs Boson can decay into various particle types, in this study, the focus neutrino pair final state:

$$h \to ZZ^* \to 2\nu + 2\overline{\nu} \tag{2.14}$$

the only IHD, this decay is very rare with less than one to one thousand chance (0.1%) (17). In the SM, the BR of the IDH is of approximately 0.1%, however, the LHC was only able to achieve a upper limit of 11% at 95% confidence level (2). The main goals of future colliders is to achieve 1% or better precision in the measurement of Higgs boson coupling (1), as BSM physics deviations form the SM are typically no larger than a few percent and any deviation from the SM mighty indicate BSM. Similarly, rares decays like $h \to Z\gamma$ that has an estimated BR of 0.15%, $h \to J/\psi$ that has a BR of 2.3×10^{-6}%, $h \to Z\psi(2S)$ with a BR of 1.7×10^{-6}% and many other rare decays (24). These rare Higgs decays will require higher precision will benefit from the improved detection precision (20).

2.3.1 The Invisible Higgs Decay

The cross-section for defines the probability of an event to happen, the cross-section for $e^-e^+ \to Zh$ and the BRs of the Higgs boson at $\sqrt{s} = 250\,\text{GeV}$ found at Figures 7 and 8 respectively, can be observed directly without reference to the Higgs decay mode (1). However h doesn't decay directly into the neutrino pair, the decay chain involves three decays. Step one is the *Higgstrahlung*:

$$e^+e^- \to Z^* \to Zh \tag{2.15}$$

an electron–positron pair e^+e^- annihilates to produce a off-shell Z^*, of which it decays into a Z boson and a Higgs boson h. The second steps split between the Z and h decay:

$$Z \to \mu^+\mu^-, \tag{2.16}$$
$$h \to ZZ^*. \tag{2.17}$$

the Z decays to a muon pair of opposite charge $\mu^+\mu^-$, and the Higgs decays into the ZZ^* pair. Then, these two decays occur at the same stage of the decay chain. Lastly, the last step is the decay of each Z and Z^* into a neutrino-anti-neutrino pair:

$$Z \to \nu\bar{\nu}, \tag{2.18}$$
$$Z^* \to \nu\bar{\nu}. \tag{2.19}$$

the combination of the three steps, leads to:

$$e^+e^- \to Zh \to \mu^+\mu^- + 4\nu, \tag{2.20}$$

This is the IHD, this decay chain leads to two relevant BR, BR($h \to$ inv) and BR($Z \to \mu^+\mu^-$). Luckly it's possible to see that actually, the decay isn't completely invisible, due to the presence of μ decayed from Z the invisibility of the neutrinos can be inferred from

the extra μ in the detectors, it's also important to note that the secondary decay can also happen with $h \to W^{\pm *}W^{\mp}$, the third highest muon pair producing event Table 6, the Higgs decay into vector boson pair, one real and one virtual, serves as a test to the electroweak gauge coupling of the Higgs (12). Additionally, the lack of neutrinos detectors in the ILC, the neutrinos produced could be any of them, however, due to the fact that the electron weakly couples with Higgs (9), it's most likely that the neutrinos are either *Tau* or *Muon* due to neutrino oscillations, the phenomenon responsible for neutrinos changing flavors.

The present analysis assumes the SM prediction for invisible Higgs decays, which proceed primarily via $h \to ZZ^* \to 2\nu + 2\bar{\nu}$, yielding a four-neutrino final state. In contrast, many BSM scenarios predict enhanced invisible decay rates, where the Higgs boson may couple to dark matter candidates, weakly interacting massive particles, or long-lived neutral states that escape detection (2, 5, 1). Any deviations of the expected SM value of the IHD branching ratio might indicated the presence of BSM, highlighting the importance of studying it.

3 The International Linear Collider

High-energy particle physics experiments can be broadly classified into two categories: fixed-target experiments, in which a beam of particles is directed onto a stationary target, and collider experiments, where two particle beams traveling in opposite directions are brought into collision. Modern discoveries at the energy frontier are predominantly driven by collider experiments, as they allow the full center-of-mass energy to be available for particle production.

Particle accelerators represent some of the most complex scientific instruments ever constructed. Advancing the frontiers of physics requires not only higher energies and luminosities, but also the development of entirely new technologies. Cutting-edge experiments cannot be built solely from pre-existing components; instead, they demand innovations in accelerator design, detector technology, data acquisition systems, and computing infrastructure.

The International Linear Collider (ILC) is a proposed electron–positron linear collider designed to operate initially at a center-of-mass energy of $\sqrt{s} = 250\,\text{GeV}$, with upgrade paths to higher energies. The ILC features two complementary detector concepts: the International Large Detector (ILD) and the Silicon Detector (SiD). These detectors are optimized for precision measurements, particularly in Higgs boson and electroweak physics.

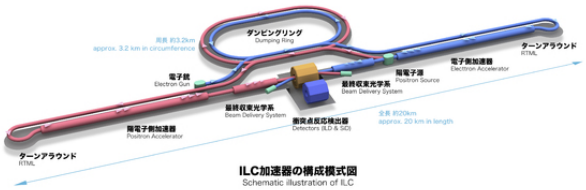


Figure 9 – Schematic illustration of the ILC. Figure taken from the ILC official website (10).

Particle accelerators can also be classified according to their geometry into circular and linear machines. Each approach has distinct advantages and technical challenges.

In circular colliders, such as the *Large Hadron Collider (LHC)*, particle bunches circulate for extended periods and can collide multiple times. However, charged particles

moving along curved trajectories emit synchrotron radiation, leading to continuous energy losses. While this effect is manageable for heavy particles such as protons, it becomes prohibitive for light particles such as electrons at high energies. The strength of the bending magnets, currently limited to approximately 20 TeV with existing superconducting technology, ultimately constrains the achievable collision energy. Increasing the energy therefore requires either stronger magnets or a larger ring circumference, as foreseen in future projects such as the Future Circular Collider (FCC).

In contrast, linear accelerators (linacs) avoid synchrotron radiation losses associated with circular motion, since the beams travel in straight lines and collide only once. Although this implies that beam reuse is not possible, linear colliders offer a cleaner experimental environment and are particularly well suited for precision measurements in electron–positron collisions. A key advantage of the ILC design is its scalability: the main linac can be extended to reach higher energies, or upgraded with more advanced acceleration technologies as they become available (1). Even within its baseline configuration, the ILC is designed to operate efficiently across a broad range of center-of-mass energies (11).

The Linear Collider Collaboration project has it's first stage planned at the 250 GeV, as dividing the project in stages would significantly lower the cost and development of the ILC. A major difference between the LHC and ILC is the improved precision for the study of the Higgs boson, the ILC detectors will benefit from 30 years of advancements in detector techology, an expected diference of more than an order of magnitude in momentum and impact resolutions, 2 times improvement in the jet energy scale and also tightened hermeticicty (1).

If individually Higgs couplings can be measured with high precision, information on properties of new physics models can be obtained. Physics at 250 GeV is subject to very small background (1). The principle of energy conservation dictate the total energy of colliding particles must be at least twice the mass of the Higgs boson.

The focus of this study is the precision Higgs couplings that happen at a center-of-mass energy of 250 GeV, the indirectly decay of Higgs into neutrinos. As mentioned in chapter 2, the decay is rightfully named invisible because it's a challenge to detect (BR = 0.1%) and the current detector technology on the LHC doesn't have enough precision to decern it. Whereas, the ILD is planned to have enough precision in their detection that missing energy of the Higgs decay no longer will an uncertainty, the invisible decay will be measurable in the same way of putting all the pieces together of a puzzle and seeing that a single piece is missing.

Energy	Reaction	Physics Goal	Pol.
91 GeV	$e^+e^- \to Z$	ultra-precision electroweak	A
160 GeV	$e^+e^- \to WW$	ultra-precision W mass	H
250 GeV	$e^+e^- \to Zh$	precision Higgs couplings	H
	$e^+e^- \to t\bar{t}$	top quark mass and couplings	A
350–400 GeV	$e^+e^- \to WW$	precision W couplings	H
	$e^+e^- \to \nu\bar{\nu}h$	precision Higgs couplings	L
	$e^+e^- \to f\bar{f}$	precision search for Z'	A
	$e^+e^- \to t\bar{t}h$	Higgs coupling to top	H
500 GeV	$e^+e^- \to Zhh$	Higgs self-coupling	H
	$e^+e^- \to \tilde{\chi}\tilde{\chi}$	search for supersymmetry	B
	$e^+e^- \to Ah, h^+h^-$	search for extended Higgs states	B
	$e^+e^- \to \nu\bar{\nu}hh$	Higgs self-coupling	L
	$e^+e^- \to \nu\bar{\nu}VV$	composite Higgs sector	L
700–1000 GeV	$e^+e^- \to \nu\bar{\nu}t\bar{t}$	composite Higgs and top	L
	$e^+e^- \to t\bar{t}^*$	search for supersymmetry	B

Table 3 – Major physics processes to be studied at the ILC, together with the lowest center-of-mass energy at which they can be studied. Shown are the reaction, the process to be studied, and a key indicating which polarization scenario would be most advantageous (11).

3.1 The International Linear Collider

The precision Higgs program will start at $\sqrt{s} = 250\,\text{GeV}$ with the *Higgstrahlung* process, $e^+e^- \to Zh$. The precision mass measurement of the Higgs boson can be made particularly cleanly in the process $e^+e^- \to Zh$, with $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$. The distribution of the invariant mass recoiling against the reconstructed Z provides a precise measurement of m_h, that is independent of the Higgs decay mode. The advantage of the $\mu^+\mu^-X$ final state is higher precision than the e^+e^-X channel, as it suffers from larger experimental uncertainties due to *bremsstrahlung*. It should be noted that the capability to precisely reconstruct the recoil mass distribution from $Z \to \mu^+\mu^-$ is what defines the momentum resolution requirement for the ILC detector.

The combination of the recoil mass distribution and the precise measurement of the $\mu^+\mu^-X$ channel is specially relevant for the study of the IHD, as these measurements only use the information from the leptonic decay products of the Z and are independent of the Higgs decay mode, this technique is powerful as it allows to determine the absolute BR of the IHD (8).

The next generation of particles colliders like the ILC requires a new generation of collider detectors to study the electroweak symmetry breaking and potential discoveries beyond the SM. The SiD is the detector for the ILC, with many years of accumulated

research by physicists and engineers, the SiD will be equipped with systems capable of excellent measurements of jet energies, based on the Particle Flow Algorithm (PFA), an algorithm by which the reconstruction of both charged and neutral particles is accomplished by an optimized combination of tracking and calorimetry (11). Figure 10 displays the components of the SiD, as a compact detector it's based on a powerful silicon pixel vertex detector, silicon tracking, silicon-tungsten electromagnetic calorimetry (ECAL) and highly segmented hadronic calorimetry (HCAL).

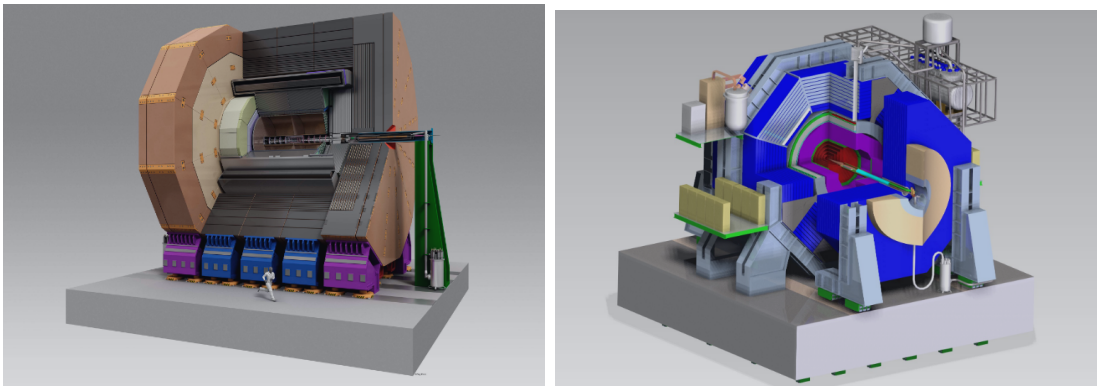


Figure 10 – Left: View of the ILD detector concept. Figure taken from (11). Right: SiD on its platform, showing tracking (red), ECAL (green), HCAL (violet) and flux return (blues). Figure taken from (11).

3.2 Cross-Section And Luminosity

The cross-section (σ) is a quantitative measure of the probability that a specific interaction occurs when two particles bunches collide with each other, it represents an effective "target area" that one particles presents to another for a given reaction. In collider experiments the flux (ϕ) is the number of collision per unit of time and surface, this value is better know as luminosity ($\mathcal{L}$) and it's usually presented in the unit of inverse barns per second $b^{-1}s^{-1}$. Both concepts relate to each other to the rate of an event X, N_X is given by:

$$\dot{N}_X = \phi\sigma_X \tag{3.1}$$

isolating σ_X and integrating over the time:

$$\sigma_X = \frac{N_X}{\int \phi dt} = \frac{N_X}{\mathcal{L}_{\text{int}}} \tag{3.2}$$

where $\int \phi dt = \mathcal{L}_{\text{int}}$ is the *integrated luminosity*, this formula shows the experimental relation between the cross-section, luminosity and the total number of events.

3.3 Components

The components of an accelerator can be divided in a few sections, the production of the particles, the acceleration and the detection. When it comes to detection, the process can be resumed into Figure 11. The tracker, calorimetry and muons system are

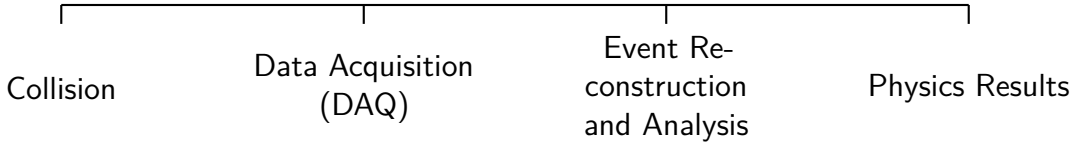


Figure 11 – Conceptual timeline of data processing in a particle physics experiment.

capable of detecting and measuring all the detector-stable particles, such as: electrons, photons, muons and a handful of hadrons. Neutrinos are not detector-stables particles due to having a incredibly weak interaction with matter and also requiring an specific type of detector. Neutrino detectors comes in the form of big tanks of a pure substance, the first experiment made to detect neutrinos had a tank full of chlorite 17 ${}^{37}_{17}\mathrm{Cl}$ (25), an example of neutrino interaction is:

$$\nu_e + n \rightarrow p + e^- \tag{3.3}$$

where ν_e is the electron-neutrino, n is a neutron, p a proton and e^- a electron, also the initial and. This neutrino interaction changes a neutron to a proton but what actually happened is change from a d down quark to a u up quark:

$$\nu_e + d \rightarrow u + e^- \tag{3.4}$$

the total charge of the system didn't change and the other two quarks present in the neutron remain unaffected due to the short range of the weak force. This meant that inside the chlorite tank the following reaction happened:

$${}^{37}_{17}\mathrm{Cl} + \nu_e \longrightarrow {}^{37}_{18}\mathrm{Ar} + e^- \tag{3.5}$$

the presence of argon 18 ${}^{37}_{18}\mathrm{Ar}$ meant that a neutrino interaction occurred. For colliders like the LHC, FCC and ILC to have neutrinos colliders, it would require the construction of big underground tank just for neutrinos and the costs of these implementations are simply beyond the budget and necessity of any of these programs.

3.3.1 Data Acquisition

The collision of e^-e^+ particle bunches results in an extremely high event rate, with an order of magnitude of $10^9 \mathrm{events}\backslash s$ (26), posing several challenges for data acquisition and analysis:

1. Proton–proton collisions occur at very high frequencies, producing vast volumes of raw data.

2. Limitations in trigger bandwidth and storage capacity allow only a fraction of these events to be recorded.

3. Of the recorded events, only a small subset corresponds to processes relevant to the physics goals of the experiment.

not only is physically impossible to collect data from all *pp* collisions, but also, it's beyond necessity to collect them all. To measure only "interesting events", a mechanism called trigger is introduced. The purpose of triggers is to only allow the detection and selection of events given set parameters.

Interaction point proton beams, tracking system, electromagnetic (EM) calorimeter, hadronic calorimeter and muon system are the components presents in particles colliders, individually they all have specific purposes and the combination of the data produced by each allows the recreation of decay chains. Therefore, in the next sections, a summary of how event detection and selection and how problems such as the millions of particles interactions are tackled will be shown.

3.3.2 Calorimeter

The purpose of the calorimeter is a detector whose purpose is to fully absorb the particle and identify them, in another words, it measures charged and neutral particles. Due to the inherently statistical nature of the measurement process, it's not possible to select and measure an individual, predetermined particle, only its type, the ILD will be equipped with calorimeters for a variety of particles, but still, a particle physics experiment has millions/thousands of collisions that produces even more particles and secondary particles, particles that results from the decay of the primary particles. The calorimeter cannot selectively register only the particles originating from a single interaction; rather, it measures the overall rate at which such interactions occur. Consequently, it is highly improbable that all products of an individual collision are recorded in a single event.

The introduction of PFAs in previous detectors, such as CDF, ZEUS have resulted in significant improvements of the jet energy resolution compared to methods based on calorimetric measurement alone. For the SiD, a PFA approach is necessary to reach the goal of obtaining a measurement uncertainty on the jet energy resolution of the order of 3% or better for jet energies above 100 GeV (11), especially relevant for the IHD.

3.3.3 Vertex

When particles undergo collisions, additional particles can be produced. The spatial points at which these newly created particles originate are referred to as primary vertices, whereas the locations at which they subsequently decay are termed secondary vertices (27).

3.3.4 Tracking

Tracking detectors are designed to measure the trajectories of charged particles through their ionization energy loss in the detector material. As a consequence, only electrically charged particles leave direct signals in the tracking system.

The combination of tracking information with energy deposits in the calorimeters allows for an efficient and nearly unambiguous identification of the main particle species produced in high-energy collisions and the tracking system is key ILC physics goal, imposing performance requirements on tracking system that exceed those met by any previous system, improvements in the momentum resolution over a broad p_T spectrum (11), especially considering the p_T is used as an selection criteria in the study of the IHD (3, 2).

The silicon-based highly efficient charged particle tracking is achieved by the SiD using the pixel detector and main tracker to recognize and measure prompt tracks, in conjunction with the ECAL (11). Every detector-stable particle has a characteristic pattern in calorimeter detection:

- $e^{\pm}$: A reconstructed charged track associated with a compact electromagnetic energy deposit in the electromagnetic calorimeter.
- γ: No reconstructed track, accompanied by an electromagnetic energy deposit in the calorimeter.
- $\mu^{\pm}$: A charged track with minimal energy deposition in the calorimeters and a characteristic penetrating behavior through the detector.
- $\tau^{\pm}$: Identified indirectly through their decay products, which may include charged tracks and calorimetric energy deposits.
- Hadrons: Charged hadrons produce tracks accompanied by energy deposits in the calorimeters, while neutral hadrons are observed only via calorimetric signatures.
- Quarks: Not observed directly, but manifest experimentally as collimated jets of hadrons resulting from the hadronization process.

Out of all the particles in the list, only muons $\mu^{\pm}$ are relevant to the invisible decay. It's important to note that particle identification relies on the complementary information provided by the tracking system and the calorimeters, exploiting differences in charge, mass, and interaction mechanisms.

3.3.5 Muon Chambers

In detectors like the LHC, it possesses special detectors just for muons, the Muon Chambers. Muons are charged particles like electrons and positrons, however, they are 200 times heavier and with that muons can penetrate several meters going beyond the standard calorimeters. Therefore, Muons require specialized detectors.

3.4 Reconstruction

A *pp* collision, that produces a particle(s) and that subsequently decays into more particles, can be seen as puzzle. Such puzzle is always incomplete because of the statistical improbability of detecting all the particles involved in a single decay chain, with data collected in the detectors, sufficient produced collisions, pieces can be build up to form a sufficient picture for the extraction of information.

3.4.1 LCIO

The LCIO software package provides a Event Data Model (EDM), the EDM is also integrated into Key4hep, the background MC files are `slcio` files and Key4hep offer tools to temporarily convert into `edm4hep.root` files during the simulation process. The conversion process requires a mapping, to properly translate the collections between file type, this allows to keep consistency with the signal files. The purpose of EDM is to have a common file format for Data Analysis. Key4hep does allows a permanent conversion with the `lcio2edm4hep` command but due to the heavy and extensive amount of `slcio` (28) files, the space required for all the generated events was beyond the accessible for this experiment.

From the DESY network. The available `slcio` files used in this analysis can be grouped into six main categories according to their underlying physics processes as described in Table 4. The Inclusive Higgs had to be excluded from the final simulated events since the Inclusive and Exclusive channels contain the same events, choosing Exclusive over Inclusive was due the fact it offered better event distinction. This classification facilitates a systematic definition of signal and background samples and simplifies the organization of the analysis workflow.

Table 4 – Classification of simulated processes used in the analysis.

Category	Description
Inclusive Higgs production	Higgs production processes without explicit constraints on the Higgs decay mode at generator level.
Exclusive Higgs production	Higgs production processes in which the Higgs decay channel is explicitly specified at generator level.
Two-fermion processes	Processes leading primarily to dilepton or quark–antiquark final states.
Four-fermion (boson pair)	Final states originating from the decay of a pair of electroweak gauge bosons (WW or ZZ).
Four-fermion (single boson)	Processes in which the final state arises from the decay of a single W or Z boson.
Other multi-fermion processes	Processes with three-, five-, or six-fermion final states not covered by the previous categories.

3.4.1.1 File Selection

There are 449 possible events in the $\sqrt{s} = 250\,\text{GeV}$ energy range that are available to be simulated from the collision, each event has an unique `process_id`, a six digit number `XXXXXX`, and are split in multiple `slcio` files, resulting in more than 130k total `slcio` files. The goal is to only study the left polarized electrons and right polarized positorns, $e_L^- e_R^+$, since it's expected for the ILC to reach the +80% electron, $P(e^-) = -0.8$ and = 30% positron, $P(e^+) = +0.3$ beam polarization at $\sqrt{s} = 250\,\text{GeV}$. Therefore, filtering to only choose `eL.pR` files from the available total of 16 beam polarizations, reduces the total events to 122 from the original 449, Table 5 has listed all types of polarizations available.

The choice of limiting the analysis to only $e_L^- e_R^+$, was not only a constrain within the scope of this thesis, but, it also coincides with the improved precision in the polarization of particles bunches in the ILC. A algorithm was used to collect all the files and added as the Key4Hep input for each `process_id` separately, with the full list can be found at ANNEX C.

3.4.1.2 Validation and Sanity Checks

The six categories of the Inclusive and Exclusive Higgs shown on Table 7 behave similarly. Analyzing the Table 6, the highest cross-section is `e2e2h`, which represents the *Higgstrahlung* without and specific final decay:

$$\texttt{e2e2h} \equiv e^+e^- \to Zh \tag{3.6}$$

The first `e2` is a charged lepton final state, the second `e2` is also a charged lepton final state and for `h` means there is a Higgs in the event. After filtering all the available MC files, only 13 are muon-producing events, 1 Inclusive and 12 Exclusive. Experimentally,

Table 5 – Beam-polarization labels and their physics equivalents. The polarization symbols are: `L` = left-handed, `R` = right-handed, `B` = unpolarized or mixed, `W` = "wide". The row `eL.pR` (corresponding to $e^-_L e^+_R$) is highlighted because it is the only configuration used in the present study.

Label	Physics notation
`eB.pB`	$e^-_B e^+_B$
`eB.pL`	$e^-_B e^+_L$
`eB.pR`	$e^-_B e^+_R$
`eB.pW`	$e^-_B e^+_W$
`eL.pB`	$e^-_L e^+_B$
`eL.pL`	$e^-_L e^+_L$
`eL.pR`	$e^-_L e^+_R$
`eL.pW`	$e^-_L e^+_W$
`eR.pB`	$e^-_R e^+_B$
`eR.pL`	$e^-_R e^+_L$
`eR.pR`	$e^-_R e^+_R$
`eR.pW`	$e^-_R e^+_W$
`eW.pB`	$e^-_W e^+_B$
`eW.pL`	$e^-_W e^+_L$
`eW.pR`	$e^-_W e^+_R$
`eW.pW`	$e^-_W e^+_W$

the probability of and event is proportional to their cross-section (σ_i), Table 6 has the visualization and values for the cross-section respectively. Other processes `e2e2h_i` where `i` are the Exclusive *Higgsstrahlung* events with the final state specified. Therefore, from Table 6, the sum of the cross-sections of all `e2e2h_i` should be equal to `e2e2h` within an expected error:

$$\sum_i \sigma(\texttt{e2e2h\_i}) \approx \sigma(\texttt{e2e2h}) \tag{3.7}$$

This relation is expected to repeats to all Inclusive Higgs processes, Table 7 has the list of the removed processes with their equivalent physics process, Table 8 has the sum for all inclusive processes, according to equation 3.7, and consequently the difference between cross-sections. All processes had a difference between the inclusive and the exclusive sum was less then $\Delta\sigma \lesssim 10^{-3}\,\text{fb}$ except for `n23n23h` that had a difference of $\Delta\sigma = 46.52\,\text{fb}$, the sum of the exclusive processes was higher, this unexpected behavior did not affect the final generated events because the inclusive process has been excluded. For consistency sake, all cross-section values were converted into fento barn [fb].

3.5 Monte Carlo

Key4hep employs the MC simulation method, a computational approach that uses random sampling techniques to model systems governed by probabilistic processes. In HEP,

Table 6 – Inclusive and Exclusive Higgs Muon-producing-events cross-sections for the *Higgstrahlung* with $e_L^- e_R^+$ beam polarization, ordered by decreasing cross-section, values for σ where taken from Table 28. The highlighted `signal` row had its value calculated at section 3.5.2, its uncertainty it was extracted from (4).

Process	σ [fb]
`e2e2h`	$(1.6971 \pm 0.0014) \times 10^{1}$
`e2e2h_bb`	$(9.8769 \pm 0.0008) \times 10^{0}$
`e2e2h_ww`	$(3.6317 \pm 0.0003) \times 10^{0}$
`e2e2h_gg`	$(1.3899 \pm 0.0011) \times 10^{0}$
`e2e2h_e3e3`	$(1.06406 \pm 0.00087) \times 10^{0}$
`e2e2h_cc`	$(4.9045 \pm 0.0004) \times 10^{-1}$
`e2e2h_zz`	$(4.4463 \pm 0.0036) \times 10^{-1}$
`e2e2h_aa`	$(3.8523 \pm 0.3105) \times 10^{-2}$
`e2e2h_az`	$(2.5965 \pm 0.2105) \times 10^{-2}$
`signal`	$(1.0535 \pm 0.0026) \times 10^{-2}$
`e2e2h_ss`	$(4.1748 \pm 0.3406) \times 10^{-3}$
`e2e2h_e2e2`	$(3.6996 \pm 0.0306) \times 10^{-3}$
`e2e2h_dd`	$(8.8247 \pm 0.7207) \times 10^{-4}$
`e2e2h_uu`	$(2.0365 \pm 0.1707) \times 10^{-4}$

Table 7 – Inclusive Higgs production processes removed from the simulated events. These samples include all Higgs decay modes and are therefore not used in the final analysis to avoid double counting with exclusive Higgs decay samples.

`slcio` name	Physics process
`e1e1h`	$e^-e^- \to e^-e^-h$ (*Higgsstrahlung* via Z exchange)
`e2e2h`	$e^+e^- \to e^+e^-h$ (*Higgsstrahlung* / ZZ fusion)
`e3e3h`	$e^+e^- \to \mu^+\mu^-h$ (*Higgsstrahlung*)
`n1n1h`	$e^+e^- \to \nu_e\bar{\nu}_e h$ (WW fusion)
`n23n23h`	$e^+e^- \to \nu_{\mu,\tau}\bar{\nu}_{\mu,\tau}h$ (ZZ fusion)
`qqh`	$e^+e^- \to q\bar{q}h$ (*Higgsstrahlung*)

MC simulations are employed to generate synthetic collision events based on theoretical models, cross-sections, and decay BRs. These simulated events are then propagated through detector models to reproduce experimental conditions, enabling the estimation of signal efficiencies, background rates, and statistical sensitivities.

The probability distribution of this simulation is based on the theoretical predictions of the SM, at the $\sqrt{s} = 250\,\text{GeV}$, the probability distribution is well behaved. At higher energy levels 1 TeV, SM cannot describe precisely, such theoretical limitations have led to the goals of FCC and ILC being to reach energy levels higher than what the LHC can reach with the current capabilities and future upgrades.

To simulate the experiment two types of files were used, signal and background, where each individual file has a unique probability distribution.

Table 8 – Summary of the cross-sections for the Inclusive Higgs processes and their variants. The Table lists the number of variants, the original generator-level cross-section, the sum of the cross-sections of all simulated variants, and their difference.

Process	n variants	σ [fb]	$\sum\sigma$ [fb]	$\Delta\sigma$ [fb]
`e1e1h`	12	17.671	17.672	4.948×10^{-4}
`e2e2h`	12	16.971	16.971	4.752×10^{-4}
`e3e3h`	12	16.941	16.941	4.743×10^{-4}
`n1n1h`	12	60.351	60.353	1.690×10^{-3}
`n23n23h`	18	67.111	113.628	4.652×10^{1}
`qqh`	12	343.030	343.040	9.605×10^{-3}

Table 9 – Impact of different `e2e2h` process types on the dimuon invariant mass $m_{\mu\mu}$ and the recoil mass m_{recoil} in an invisible Higgs analysis. The highlighted rows are for the processes that contribute to the background samples. The `e2e2h_e3e3` and `e2e2h_e2e2` doesn't affect neither the invariant or recoil mass reconstruction, due the fact that any final state that contains more than two muons are discarded.

Process type	Extra muons	$m_{\mu\mu}$	m_{recoil}	Notes
`e2e2h`	No	None	None	Ideal signal reference
`e2e2h_bb`	No	None	None	Hadronic Higgs decay
`e2e2h_cc`	No	None	None	Hadronic Higgs decay
`e2e2h_gg`	No	None	None	Gluon final state
`e2e2h_aa`	No	None	None	Photonic Higgs decay
`e2e2h_ww`	Possible	Moderate	Moderate	Muons from W decays
`e2e2h_zz`	Possible	Moderate	Moderate	Muons from secondary Z
`e2e2h_az`	Possible	Moderate	Moderate	Mixed bosonic decay
`e2e2h_e3e3`	Yes	None	None	$\tau \to \mu$ contamination
`e2e2h_e2e2`	Yes	None	None	$h \to \mu^+\mu^-$ (4μ final state)

3.5.1 Weights

As the simulation allows the luxury of choosing the which and how many events are simulated, it's important that the generated samples of all simulated events are all proportional to their real probability, as the goal is be as close as reality. The MC method allows the introduction of weights, w_i, where i is an specific event, to make sure that events with different probabilities are properly represented. The weights are defined as:

$$w_i = \frac{\sigma_i \cdot \mathcal{L}_{\text{target}}}{N_{\text{gen}}} \tag{3.8}$$

where σ_i is the process cross-section, $\mathcal{L}_{\text{target}}$ is the target integrated luminosity (not to be mistaken with the Lagrangian) and N_{gen} the total number of events generated. At $\sqrt{s} = 250\,\text{GeV}$ a plethora of events occur along the desired decay and all the background MC files serve the purpose of simulating other events, A summary of the background files are located at: Production Summary (PS) (29). For the cross-section the values are located at ELOG gendata (ELOG) (15).

It's important to note that to get the associated cross-section σ_i for each event, the tables from PS and ELOG had to be combined, with `process_id` as the primary key, both tables have different values for the integrated luminosity $\mathcal{L}$ and N_{gen}, neither of these values where used, these are useful in the scenario where a single MC sample is analyzed. For a combined sample of MC generated events, the $\mathcal{L}$ is an arbitrary global variable and N_{gen} are dependent on the number of simulated events for each event, all events had the same number for the max number of events simulated $N_{\texttt{EvtMax}}$, later explained at Section 4.1, this number is different from the actual number of generated events N_{gen}. The values chosen were $\mathcal{L} = 250\,\text{fb}^{-1}$ and $N_{\texttt{EvtMax}} = 10^6$; with these defined proprieties, w_i is be calculated, the resulting values can be found at Figure 12 sorted in descending order.

3.5.2 Signal Weight

Unlike the background weights, that have the cross-section available at (15). The signal weight have to be calculated separately, the signal cross-section was calculated by multiplying, the *Higgstrahlung* cross-section, the BR for Higgs to invisible, BR($H \to \text{inv}$), and BR of the Z decay into a muon pair $\mu^+\mu^-$, BR($Z \to \mu^+\mu^-$). This leads to the signal cross-section σ_{signal} formula:

$$\sigma_{\text{signal}} = \sigma_{Zh} \times \text{BR}(h \to \text{inv}) \times \text{BR}(Z \to \mu^+\mu^-) \tag{3.9}$$

the values for the *Higgstrahlung* cross-section is dependent on the beam polarization, Table 10 has the *Higgstrahlung* cross-section values for the ILC program.

Table 10 – SM *Higgstrahlung* cross-section at $\sqrt{s} = 250\,\text{GeV}$ for unpolarized and polarized beams $P(e^-) = -0.8$, $P(e^+) = +0.3$ (2). Full table at ANNEX A.

Center-of-Mass Energy [GeV]	Unpolarized [fb]	Polarized [fb]
250	211	313

To study the IHD the parameters used were: left polarized beams of the ILC, $\sigma(e^+e^- \to Zh) = 313\,\text{fb}$, BR($h \to \text{inv}$) = 0.1% (2, 5) and BR($Z \to \mu^+\mu^-$) = 3.3662(66)% (14). Substituting the values, results in:

$$\sigma_{\text{signal}} = 313 \times 10^{-3} \times 3.37 \times 10^{-2} = 1.05 \times 10^{-2}\,\text{fb} \tag{3.10}$$

now this can be substituted in equation 3.8.

3.5.3 Error Estimation

Simulating an experiment has two important aspects. Firstly, unlike real experiments, computers work in a deterministic way, a program will always produce the same results when ran and randomness has to be written into a program, computers are not

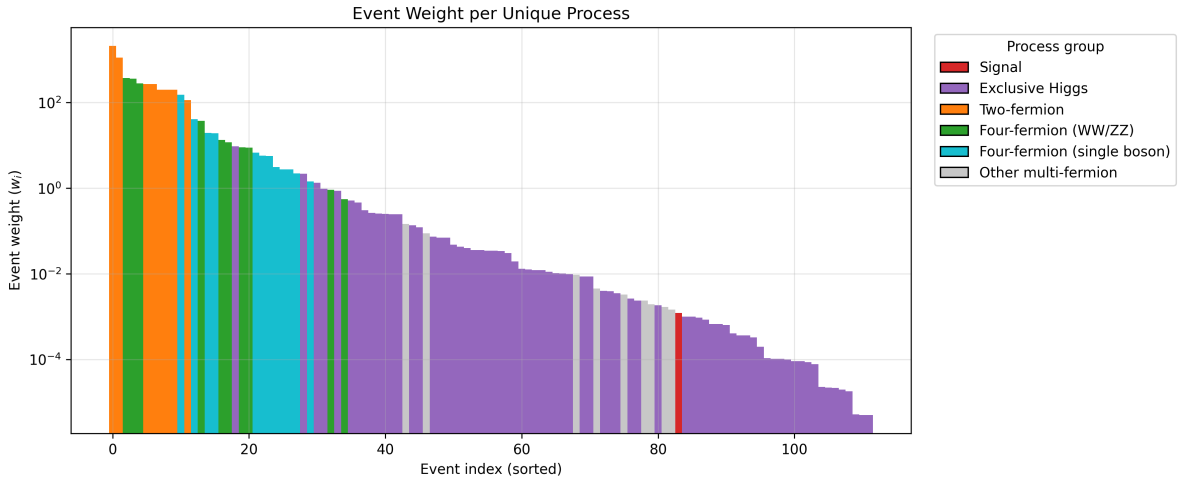


Figure 12 – Event weights w_i for each process in *log* scale, sorted from highest to lowest, colored based on categories from Table 4, with Signal (Red).

actually capable of real randomness, they are in fact pseudo-random[1]. Pseudo-randomness has the important aspect that is reproducibility, allowing to choose initial seeds that define all random future values. Secondly, simulations are build on what it's understood about reality, not how it really is, therefore, simulated errors will not provide a complete picture of how real experiments will behave. All real experiments have uncertainty in them which can be divided into two categories:

- Systematic, which is due to imperfections in the equipment, improper or biased observation and the presence of additional physical effects not taken into account (For the collision energy of $\sqrt{s} = 250\,\text{GeV}$, the expected results are well described within SM).

- Random, uncertainties that are an unavoidable result of measurement, regardless of how well tools are calibrated.

Both types of uncertainty are not present in computer simulations[2], but because they are so intrinsic to reality, that also simulating uncertainty can and should be done. As the ILC program evolves, systematic errors such as the precision in the integrated luminosity $\mathcal{L}$, beam polarization has 0.25% and the center-of-mass energy $\sqrt{s}$ has 0.01% expected precision (2), trigger efficiency and many more systematic errors that can be measured as soon they are developed and tested.

[1] True or simulated randomness is an extensive field in computer science, Quantum Computers can be used to get real randomness since they are nondeterministic.

[2] Different results can happen because of hardware defects or external source, e.g. Cosmic Rays can cause the bits in a computer to flip. But such problems are so exceptional that it can be ignored.

Table 11 – Systematic uncertainties included in the projected sensitivity for ILC configuration. Values found at (2, 11, 4).

Category	Source	Assumed Uncertainty
Normalization		
Integrated luminosity	$\mathcal{L}$	0.25%
Center-of-Mass Energy	$\sqrt{s}$	0.01%
Beam polarization	$P(e^{\pm})$	$0.2 - 0.5\%$
Signal cross-section	$\sigma(Zh)$	$1 - 2.5\%$
Background normalization	σ_{bkg}	$5 - 10\%$
Detector / Reconstruction		
Muon momentum scale	p_T^{μ} scale	0.1%
Muon momentum resolution	smearing	$2 - 5 \times 10^{-5}$ GeV
Selection efficiency	ϵ_{sel}	1–2%
Theory / Modeling		
ISR modeling	structure functions	2%
Beamstrahlung modeling	beam spectrum	1%
Generator dependence	dominant backgrounds	0.5%

3.5.4 Histograms

The Monte Carlo Simulation possess the theoretical tools for error propagation. the weighted mean is:

$$\hat{X} = \sum_{i}^{N} w_i X_i \tag{3.11}$$

Due the fact that the weights can distort how much "information" the simulated events effectively have, considering that events weights values are within a few orders of magnitude, varying from $w \in [2.10 \times 10^3, 5 \times 10^{-6}]$. For a weighted histogram, the weight associated with each bin is:

$$W_{\text{bin}} = \sum_{i} w_i X_i \tag{3.12}$$

statistical uncertainty in a bin is computed as:

$$\sigma_{\text{bin}} = \sqrt{\sum_{i} w_i^2} \tag{3.13}$$

then σ_{bin} is plotted on top of each bin in the histogram[3].

3.5.4.1 Signal Error

The first step in calculating the uncertainty in the IHD Cross-Section is to divide between the systematic and statistical uncertainties:

$$\delta\sigma_{\text{signal}} = \sqrt{(\delta\sigma_{\text{syst}})^2 + (\delta\sigma_{\text{stat}})^2} \tag{3.14}$$

[3] All these formulas where taken from <https://www.zeuthen.desy.de/~wischnew/amanda/discussion/wgterror/working.html>.

where $\delta\sigma_{\text{syst}}$ is the systematic uncertainty and $\delta\sigma_{\text{stat}}$ the statistical uncertainty that is based on the MC simulation. The formula 3.10 is composed of the multiplication of three values, the uncertainty in σ_{signal} is calculated with the assumption that the values are uncorrelated. The formula for the relative uncertainty $\delta(\sigma_{\text{signal}})$ is limited to the uncertainty in the BR($Z \to \mu^+\mu^-$):

$$\left(\frac{\delta\sigma_{\text{sys}}}{\sigma_{\text{signal}}}\right)^2 = \left(\frac{\delta\sigma(Zh)}{\sigma(Zh)}\right)^2 + \left(\frac{\delta\text{BR}(h \to \text{inv})}{\text{BR}(h \to \text{inv})}\right)^2 + \left(\frac{\delta\text{BR}(Z \to \mu^+\mu^-)}{\text{BR}(Z \to \mu^+\mu^-)}\right)^2 \tag{3.15}$$

where $\delta(x)$ is the uncertainty of a variable x, the projected uncertainty for the δBR($h \to$ inv) is set to zero since the goal is project the limit of the BR. Now the value can be substituted into the weight formula 3.8:

$$w_{\text{signal}} = \frac{\sigma_{\text{signal}} \times \mathcal{L}}{N_{\text{gen}}} \tag{3.16}$$

The expected weighted signal yield is:

$$W_{\text{signal}} = \sum_i w_i \tag{3.17}$$

where i is for a given set signal samples, the number of samples will depend on how many samples remain after all the cuts are done. Then the statistical uncertainty is:

$$\sigma_{\text{stat}} = \sqrt{\sum_i w_i^2} \tag{3.18}$$

since this is the sum over only signal events and the weight is equal for all values:

$$\sigma_{\text{stat}} = \sqrt{N_{\text{sig}} w^2} = \sqrt{N_{\text{sig}}} \cdot w \tag{3.19}$$

resulting the total cross-section uncertainty of:

$$\delta\sigma_{\text{signal}} = \sqrt{(\delta\sigma_{Zh})^2 + (\delta\sigma_{\text{BRZ}})^2 + (\delta\sigma_{\text{stat}})^2} \tag{3.20}$$

$$\tag{3.21}$$

dividing by σ_{signal} results in the relative uncertainty. This uncertainty estimate is used solely as an input to the counting-based sensitivity study presented later. A full systematic treatment including correlated nuisance parameters is beyond the scope of this work.

4 Simulation and Data Analysis

4.1 Key4hep Python Steering File

To simulate there's a series of files needed to use the Key4hep framework. Gaudi is well established software framework used to create algorithms which process HEP events and scripts which execute a chain of such algorithms on the data (30), used in experiments like LHCb, ATLAS and now Key4hep. to process save the data into a ROOT file. The entirety of the code is located the author's github[1]. Gaudi algorithms are implemented in `C++`, and are divided into several categories, Table 12. The `MConsumerAlg` is consumer

Table 12 – Types of Gaudi algorithms in the Gaudi/Key4hep framework.

Algorithm Type	Role / Description
`Producer`	Produces one or more outputs without requiring inputs; typically used for constant data or I/O tasks.
`Consumer`	Takes one or more inputs without producing outputs; used to consume data from the event store.
`Transformer`	Takes one or more inputs and produces one output; used for data transformations.
`MultiTransformer`	Takes one or more inputs and produces multiple outputs; generalised transform algorithms.
`MergingTransformer`	Combines identical inputs into a single output; useful for merging collections.
`SplittingTransformer`	Takes one input and produces multiple identical outputs; for splitting collections.
`FilterPredicate`	Evaluates conditions on input data, returning true/-false; used to gate algorithm execution.
`ScalarTransformer`	Converts scalar transformations into vectorized ones using 1-to-1 mapping.

type Gaudi algorithm was designed to process and select all events based on the presence of two muons. A steering `python` script serves as an interface between the user and the Gaudi functional implementation, integrating the Gaudi algorithm into the Key4hep application, the `ApplicationMgr` has three important options:

- `TopAlg`: The list of algorithms, like `MConsumerAlg`,
- `EvtMax`: Number of simulated events, $N_{\texttt{EvtMax}} = 10^6$,
- `ExtSvc`: External Services, Marlins Processors, `Lcio2EDM4hepTool`

[1] <https://github.com/mazeeqe/tomaz-example>

The choice of total maximum simulated events was $N_{\texttt{EvtMax}} = 10^6$ with the intent of not missing any possible generated events. Since high weight values means that certain events are under-represented, some events had the requirement of a large value of $N_{\texttt{EvtMax}}$, to generate as much events N_{gen} as possible, to be physical consistent with SM predictions. The number of $N_{\texttt{EvtMax}}$ is not the same as the number of generated events N_{gen}.

The external services, Marlins Processors, `lcio_writer` and `Lcio2EDM4hepTool` were needed for the background simulation, these external services convert `slcio` files into a `EDM` format. A detailed workflow with code examples can be found at the ANNEX A.

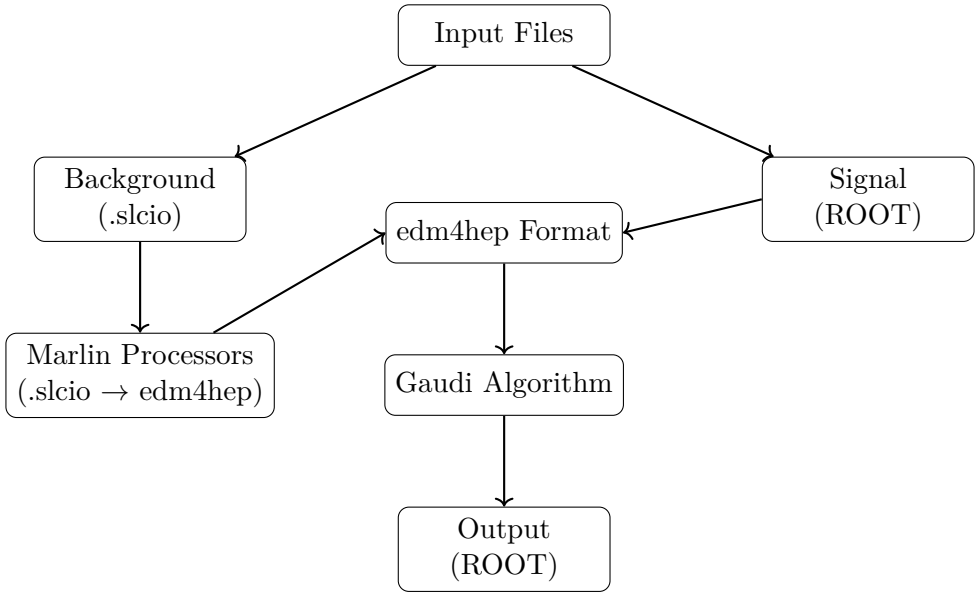


Figure 13 – Input files processing workflow diagram. Source: Author's own creation.

4.1.1 Input Selection

With the expected necessity of the MC weight calculation, simulation runs where separated for the signal and each `process_id`. As described in Section 3.4.1.1, only files that had `eL.pR` on their name where valid as input files. The selection algorithm has the `process_id` and the path for the `250-SetA` as inputs, a regex filtered to only add files with `eL.pR` and `process_id` in their name, then all files found where added to list. Finally, the list was used as an input for `io_svc.Input`.

4.1.2 Output Data

The simulation produces an plethora of data, the data is stored at `collections` portion of the code. The `PandoraPFOs` is used to obtain the data from the reconstructed particles. The selected particles can be μ or $\bar{\mu}$, particles are selected based on their `PDG ID` with the anti-particle being the negative value of the matter ID, for $\texttt{ID}(\mu) = 13$ and $\texttt{ID}(\bar{\mu}) = -13$, so the algorithm selected all particles that had $\texttt{ID} = |13|$.

The ROOT framework is used to create and save the desired variables, the `TFile()` function is used to create the ROOT file, the `TTree()` function creates a tree and within a tree there are `Branch` for which they are used to store data for each variable. The variables for the event system used for the cutflow Section 4.4 are present in Table 13, other variables were produced and used during the simulation runs, a complete list of stored analysis variables is provided in Appendix A.

Table 13 – List of variables used for the cutflow, grouped by category.

Category	Variable	Description
Global event	`nCharged`	N_{ch}
Visible system	`visibleEnergy`	E_{vis}
	`visiblePt`	p_T^{vis}
Observables	`m_invMass`	$m_{\mu\mu}$
	`m_recoilMass`	m_{recoil}
	`m_totalEnergy`	E_{tot}
Leading Muon	`m_mu1_E`	E^{lead}
	`m_mu1_pt`	p_T^{lead}
	`m_mu1_charge`	q^{lead}
Subleading Muon	`m_mu2_E`	E^{sublead}
	`m_mu2_pt`	p_T^{sublead}
	`m_mu2_charge`	q^{sublead}
Muon pair	`m_transverse_momentum`	$p_T^{\mu\mu}$
Missing	`m_missingEnergy`	E_{miss}
	`m_met`	E_T^{miss}

After each run a ROOT file was saved, the output file was then save into its respective folder, the output folder was divided for each `process_id`, except for the signal output, where it had its proper folder. The purpose of all these variables will be explained in next sections. For the further data analysis the following global parameters where established at Table 14.

Table 14 – Global parameters used in the analysis.

Parameter	Symbol	Value	Description
Integrated luminosity	$\mathcal{L}$	$250\,\text{fb}^{-1}$	Total delivered luminosity
Maximum generated events	$N_{\texttt{EvtMax}}$	10^6	Generator event limit per sample
Center-of-mass energy	$\sqrt{s}$	250 GeV	e^+e^- collision energy
Electron polarization	e_L^-	-80%	Left-handed electron beam
Positron polarization	e_R^+	$+30\%$	Right-handed positron beam

4.2 Signal and background definition

4.2.1 Signal

The signal process considered in this analysis corresponds to the invisible decay of the SM Higgs boson produced via *Higgsstrahlung*, specifically the pair of muons produced in association with the neutrinos described in the equation 2.20. Events are generated using a collection of ROOT files `Dirac-Dst-E250-e2e2h_inv.eL.pR_bg-000XX.root`, where `XX` goes from 01 to 12, the naming convention for the signal input files names are described at 15.

Table 15 – Signal ROOT file naming convention used in the Key4hep framework.

Filename Part	Description
`Dirac`	DIRAC-based production workflow
`Dst`	Data type: DST-level output
`E250`	Center-of-mass energy $\sqrt{s} = 250\,\text{GeV}$
`e2e2h_inv`	Signal process: $e^+e^- \to Zh \to e^+e^- + \text{inv}$
`eL`	Left-handed electron beam polarization
`pR`	Right-handed positron beam polarization
`bg-000XX`	Production file index, with `XX` ranging from 01 to 12
`root`	ROOT file format
Complete signal filename	`Dirac-Dst-E250-e2e2h_inv` `.eL.pR_bg-000XX.root`

4.2.2 Background

The majority of the processes produced at an e^+e^- collider are well understood SM reactions that may mimic the experimental signature of the signal under study. These processes are collectively referred to as background events. Backgrounds can be classified into two categories: reducible and irreducible.

Reducible backgrounds correspond to processes whose final states differ from the signal topology but are nevertheless selected due to detector limitations or imperfect event reconstruction and particle identification. Such backgrounds arise from finite detector resolution, reconstruction inefficiencies, or misidentification effects. In the context of the IHD analysis, examples of reducible backgrounds include four-fermion final states and semi-leptonic processes in which missing energy originates from undetected particles or reconstruction effects.

Irreducible backgrounds are processes that exhibit the same visible final-state topology as the signal and cannot be efficiently separated on an event-by-event basis. These backgrounds differ from the signal only through their kinematic distributions and production rates, their occurrence can happen at any step of the invisible decay

chain. A prominent irreducible background for the $Zh,\ h \to$ inv search is the process $e^+e^- \to Z(\mu^+\mu^-)\nu\bar{\nu}$, which results in the same observable final state as the signal.

The first step in the decay chain also has a dominant background $e^+e^- \to WW$, as this is also a *Higgsstrahlung* but with a W boson instead of Z (2).

A succinct description of expected background processes:

1. $e^+e^- \to ZZ$ with one $Z \to \mu^+\mu^-$ and the other $Z \to \nu\nu$: This produces exactly two muons and missing energy from the neutrinos — irreducible because the final state is identical to the signal, aside from kinematics.

2. $e^+e^- \to WW \to \mu\nu\,\mu\nu$: Both W bosons decay leptonically to muons and neutrinos. This leads to two muons plus missing neutrinos. Kinematically this can differ from Zh, but it still enters the sample becayse of the muon pair.

3. $e^+e^- \to Z\nu\bar{\nu}$ with $Z \to \mu^+\mu^-$: This is a single-boson production with additional missing neutrinos from processes such as Z radiation and neutrino production. It can look like the signal if the extra neutrinos produce significant missing momentum.

4. Two-fermion processes $e^+e^- \to \mu^+\mu^-$: Pure dilepton or dijet production with initial-state radiation (ISR) or mismeasured detector effects can fake missing energy.

5. Radiative or ISR processes: Processes like $e^+e^- \to \mu^+\mu^-\gamma$ where the photon escapes detection (e.g., down the beam pipe) can fake missing energy.

4.2.2.1 Background Files

All processes not classified as signal and passing the event selection are considered background. The background files have the same naming pattern, Table 16 has a file as an example. The final section of the file `nXXX.d_dstm_YYYY_ZZ`, can have a variety of values for each `XXX`, `YYYY` and `ZZ`, but nonetheless, they all belong to the same event. Table 28 has all events simulated and how many `slcio` files belonged to each event, there is 123 different events, the sum of files for each event resulted in a total of 75751 files used as input for background simulation.

4.2.3 Data Cleaning

An intrinsic aspect of MC simulation that differs from reality is that computers are pseudo-random, meaning that there's a non-zero chance of repeated generated events, `python` has functionalities to check and remove duplicated events, for the produced samples, no duplicated events was found.

Table 16 – Breakdown of the `slcio` file naming convention.

Filename Part	Description
`rv02-02`	Reconstruction software version
`sv02-02`	Simulation software version
`mILD_l5_o1_v02`	Detector model (ILD), layout 5, option 1, version 02
`E250-SetA`	Center-of-mass energy $\sqrt{s} = 250\,\text{GeV}$, dataset Set A
`I500106`	Process or generator process identifier
`P4f_sw_sl`	Physics process: four-fermion ($4f$), s-channel, semileptonic
`eL`	Left-handed electron beam polarization
`pR`	Right-handed positron beam polarization
`n000`	Production job or sample index
`d_dstm_15176_2`	Data type (DST-M), production ID and batch number
`slcio`	SLCIO data file format
Complete File Name	`rv02-02.sv02-02.mILD_l5_o1_v02.E250-SetA.I500106.P4f_sw_sl.eL.pR.n000.d_dstm_15176_2.slcio`

4.2.4 Reproduction

The generated data is dependent on the generator input files, therefore, the total amount and the characteristics of the generated samples cannot be changed by the Key4hep steering file. This results means that the data will always be the same and consequently reproducible.

4.3 Figure of Merit

One of the main challenges of performing a precision test of the SM, is to differentiate background processes from signal ones, as often background samples mimics the signal samples in some aspects that contaminate the results. As the background and signal samples are both present, the significance of the measurement, also known as FoM, is defined as:

$$FoM = \frac{S}{\sqrt{S+B}} \tag{4.1}$$

where S is the signal samples and B the background samples. Since the events are weighted, the definition of S and B is not defined by the amount, the number of data point in each category, but the yield:

$$X = \sum_i w_i \tag{4.2}$$

where $X = S, B$ and i is the sum of all weights for either the signal and background samples. The usefulness of FoM is to indicate the best balance between reducing the background without significant loss of the signal. The FoM actually has two formulas, the second formula is useful when looking for small signals out of large background ($S \ll B$)

so that its new approximate form is:

$$FoM \sim \frac{S}{\sqrt{B}} \tag{4.3}$$

but given the amount of signal samples present, equation (4.1) is the one that was used.

For analysis of the invariant and recoil mass, a cut is applied to the simulated events, reducing the size interval for both masses to the region of interest will inevitably result in the amount of signal samples present to be smaller at the advantage of removing a significant size of background samples.

The FoM formula is only useful in events predictable in SM, as the expected signal events depends in the cross-section of the process. FoM alternatives have been proposed, the Punzi FoM is designed to be able to detect new particles (31).

It's important to note that FoM assumes purely statistical uncertainties and does not account for systematic effects or correlations between background components. Therefore, it is used only as a comparative optimization metric rather than a final discovery or exclusion criterion.

Initially all different type of background files were simulated separately, after that all of them have been combined into the background samples. Out of the totality of the initial generated samples, only $6.77 \times 10^{-5}\%$ is signal as seen in the Figure 14.

4.4 Cutflow

Introducing S_0 and B_0 as the initial signal and background yields, defined at equations 4.2, and S_f and B_f as the yield after a defined cut, the following metrics can be defined:

$$\epsilon = \frac{S_f}{S_0} \tag{4.4}$$

$$R = \frac{B_0}{B_f} \tag{4.5}$$

these metrics are, efficiency (ϵ) and rejection (R), their purpose is to measure the quality of the selection procedure. A perfect selection results in $\epsilon = 0$ and $1/R = 0$, meaning that all the background samples has been removed and only signal remains. The acceptance and efficiency are computed using the unaggregated event composition after the final selection. Aggregation of minor background contributions is applied only for visualization purposes and does not affect the physics results.

For the selection criteria, the cuts where inspired by (2) and a new cut based on the energy sum of the muon pair energy and the missing energy. The baseline for the

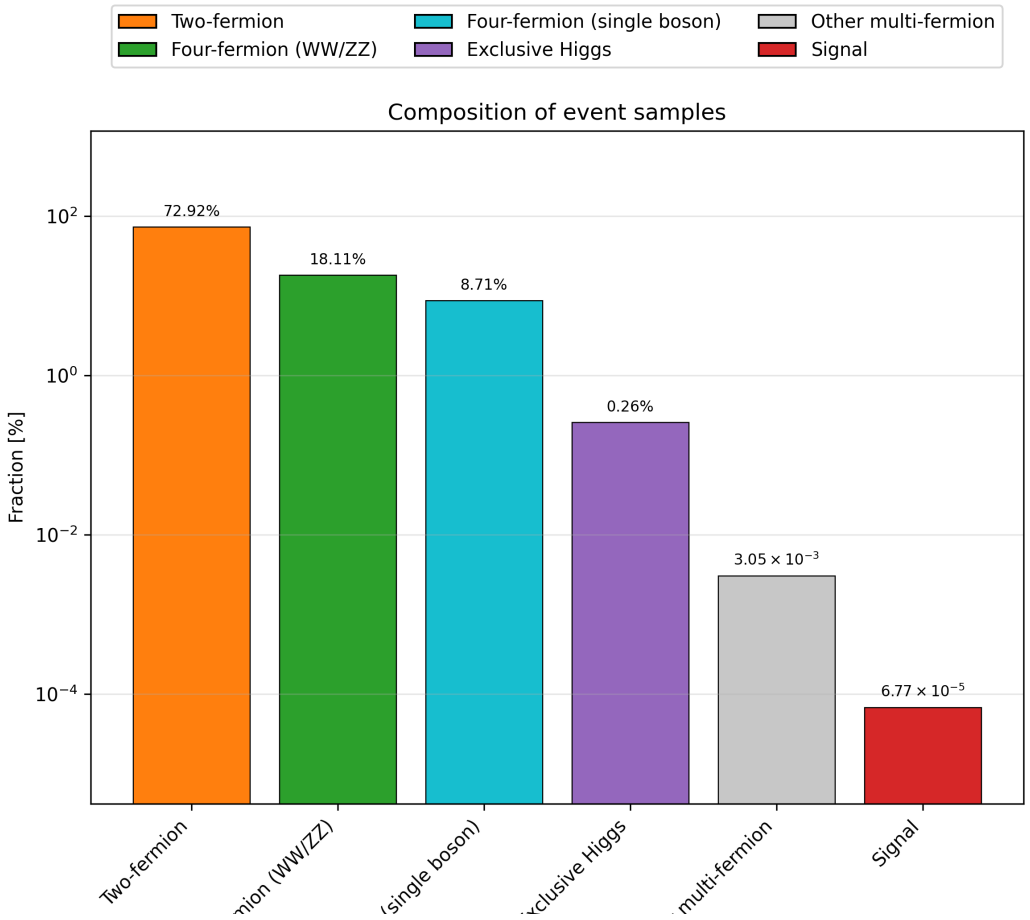


Figure 14 – Percentage composition of the uncut simulated events produced from the simulations in log scale, each bar is a category as described in Table 4. Percentage values are calculated based on the yield of each category.

selection of events was the presence of two muons μ. The IHD channel selection followed the criteria:

- All Events: all generated samples without any selection,
- Muon Pair: the presence of the dimuon, two muons of opposite charge $\mu^+\mu^-$,
- Charge Multiplicity: events where the number of charged particles is $N_{\text{ch}} \leq 4$,
- Muon Pair p_T: both muons must have $p_T > 10\,\text{GeV}$,
- Missing energy: for missing energy values higher than $E_{\text{miss}} > 40\,\text{GeV}$.
- Transverse Momentum: for values of the TM between $20 < p_T < 70$,
- Z Mass Window: for the reconstructed invariant mass values within $|m_{\mu\mu} - m_Z| < 10\,\text{GeV}$,

- h Higgs Mass Window: for the reconstructed recoil mass values within $|m_{\text{recoil}} - m_h| < 10\,\text{GeV}$.

The choice and ordering of each cut was build based on previous studies that applied the recoil mass technique and/or searched the IHD at leptonic and hadronic channels, at either the LHC (3) or the ILC (2, 4, 5, 6). Since the expected last of step of the reconstruction of the decay is the reconstruction of the recoiling Higgs, then, starting from the first cut to the last, the logic behind is to make sure that all the pieces necessary to rebuild the final state without being to restrict.

The following sections are dedicated to each cut: 4.4.1, all events, 4.4.2, multiplicity, 4.4.6, TM, 4.4.8, the invariant mass, 4.4.9, the recoil mass, describes the cuts made in order and the reason behind it. All the histograms had to have the signal contribution re-scaled due to it's small presence.

Table 17 – Signal yields S, background yields B and significance $\frac{S}{\sqrt{S+B}}$. The assumed BR is 0.1%. The integrated luminosity is $250\,\text{fb}^{-1}$.

Requirement	Signal	Background	$\frac{S}{\sqrt{S+B}}$
All Events	2.63	1.16×10^8	2.446×10^{-4}
Muon Pair	2.55	3.77×10^6	1.315×10^{-3}
$N_{\text{ch}} \leq 4$	2.50	7.60×10^5	2.863×10^{-3}
Muon $p_T > 10$	2.37	6.66×10^5	2.908×10^{-3}
$E_{\text{miss}} > 40\,\text{GeV}$	2.37	3.10×10^5	4.264×10^{-3}
$20 \leq p_T^{\mu\mu} \leq 70\,\text{GeV}$	2.20	5.14×10^4	9.720×10^{-3}
$\lvert m_{\mu\mu} - m_Z \rvert < 10\,\text{GeV}$	2.00	7374.00	0.0233
$\lvert m_{\text{recoil}} - m_h \rvert < 10\,\text{GeV}$	1.67	2079.50	0.0365

Table 18 – Signal-to-background yield ratio (S/B), the signal Efficiency (ϵ), background Rejection (R).

Requirement	S/B	ϵ	R
All Events	2.271×10^{-8}	1.000	1.000
Muon Pair	6.768×10^{-7}	0.969	0.033
$N_{\text{ch}} \leq 4$	3.284×10^{-6}	0.978	6.555×10^{-3}
Muon $p_T > 10$	3.564×10^{-6}	0.901	5.740×10^{-3}
$E_{\text{miss}} > 40\,\text{GeV}$	7.664×10^{-6}	0.901	2.670×10^{-3}
$20 \leq p_T^{\mu\mu} \leq 70\,\text{GeV}$	4.288×10^{-5}	0.837	4.432×10^{-4}
$\lvert m_{\mu\mu} - m_Z \rvert < 10\,\text{GeV}$	2.712×10^{-4}	0.759	6.359×10^{-5}
$\lvert m_{\text{recoil}} - m_h \rvert < 10\,\text{GeV}$	8.009×10^{-4}	0.632	1.793×10^{-5}

The Section 4.4.10 provides Figures for evolution of the signal and background yields for each category from Table 4, the ANNEX D has tables with the values used to plot the figures.

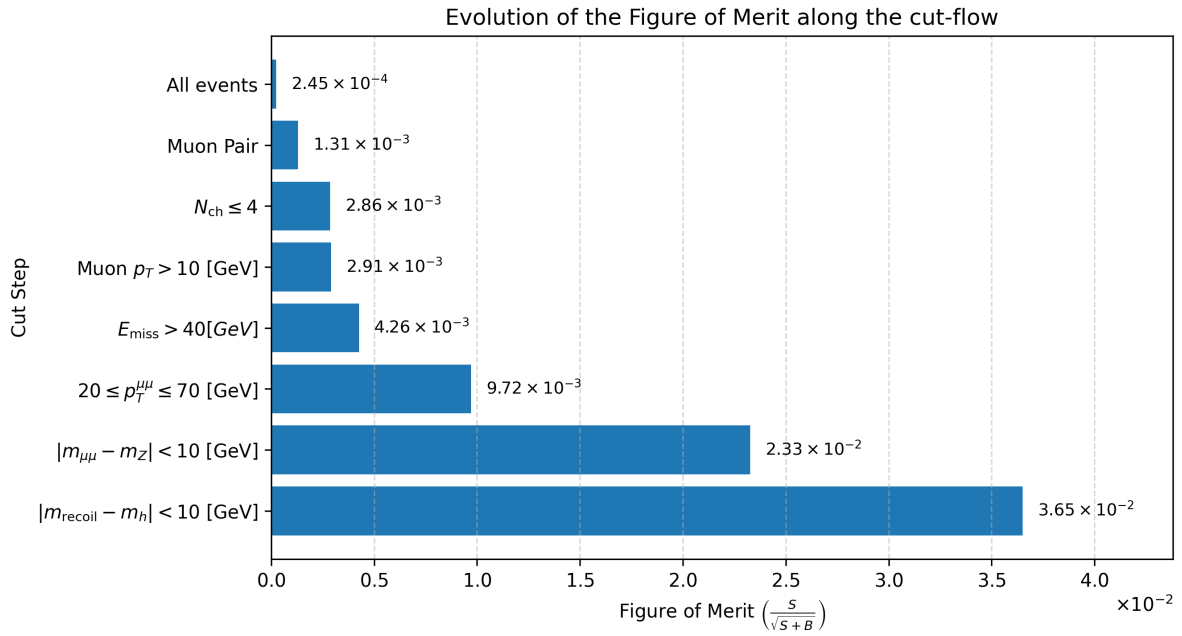


Figure 15 – Figure of Merit evolution for each subsequent cut in the cutflow. The x-axis have been normalized based on the initial FoM value with a vertical line showing starting at FoM = 1.0. All bars have annotations that displays the values for each FoM.

4.4.1 All Events

As the baseline event selection is the $Z \to \mu^+\mu^-$ channel, exactly two reconstructed muons were required in the final state. These muons were identified as the leading and subleading muons, ordered by TM. Events containing additional reconstructed muons were rejected in order to suppress backgrounds with higher lepton multiplicity. Furthermore, the final state, $Z \to \mu^+\mu^-$, also requires the two selected muons to have opposite electric charges. This requirement is expressed as either the sum of both charges be equal zero (2) or their charges multiplied are less than zero:

$$q^{\text{lead}} \times q^{\text{sublead}} < 0. \tag{4.6}$$

where q^{lead} and q^{sublead} denote the electric charges of the leading and subleading muons, respectively. This opposite-sign condition consistency effectively suppresses same-sign dilepton backgrounds arising from charge misidentification and non-resonant processes.

Studies that searches that used lepton channels limited the first selection to the number of leptons N_ℓ to be equal to two or close to it, where $\ell = e, \mu$ depends on the channel, either a pair of electrons or muons with opposite charges (3, 5, 2), other analysis also limited to the total leptonic number to be zero $n_\ell = 2$ (5), due the fact that particles of opposite charge but same flavor have opposite lepton numbers, some also used the angle between leptons as selection $\cos\theta_{\ell_+\ell_-}$ (2, 5, 3).

4.4.2 Charge Multiplicity

The charged-particle multiplicity (N_{ch}) is the number of charged particles present in the decay chain. Starting by the muon pair of opposite charge $\mu^+\mu^-$, the initial multiplicity is:

$$N_{\text{ch}}(\mu^+\mu^-) = 2 \tag{4.7}$$

to not be too strict in the selection, the selection criteria limited to four charged particles with zero total charge:

$$N_{\text{ch}} \leq 4, \tag{4.8}$$

this leaves the possibility of up to two more charged particles in the system, Figure 16 displays the initial multiplicity histogram, it's worth noting that not all signal events produced only the dimuon, this could be due to ISR, where the initial electron or position emitted particles, there's no values for $N_{\text{ch}} = 1$, since the minimum to collect data is the presence of the dimuon $\mu^+\mu^-$. A cut just for the multiplicity was a unique choice for this work, as usually this cut done in conjunction with other cuts, together with a limit to the number of leptons present (2) or when measuring the lepton TM p_T.

The charged multiplicity requirement completely suppresses the "Other multi-fermion" contribution and strongly suppresses the "Exclusive Higgs" contribution, as illustrated in Figure 30. This is consistent with the topology of the signal and dominant four-fermion backgrounds, which predominantly produce final states with $N_{\text{ch}} = 4$. Processes leading to higher charged multiplicities are therefore efficiently rejected by this selection.

The charged multiplicity distributions before and after the application of the full event selection are presented in Figures 16 and 17, respectively.

4.4.3 Muon Transverse Momentum

The TM is the magnitude of the particle momentum that it transverse to the proton beam, taking the momentum values for the x and y axis, squaring them and than adding them together, results in the TM p_T:

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{4.9}$$

The muons have been sorted by energy, the highest energy muon is named the Leading Muon (μ^{lead}) and lowest energy muon is named the Subleading Muon (μ^{sublead}). The selection criteria for the leading and subleading muons is that they must have a TM of at least $p_T > 10\,\text{GeV}$:

$$\begin{cases} p_T^{\text{lead}} & > 10\,\text{GeV}, \\ p_T^{\text{sublead}} & > 10\,\text{GeV}. \end{cases} \tag{4.10}$$

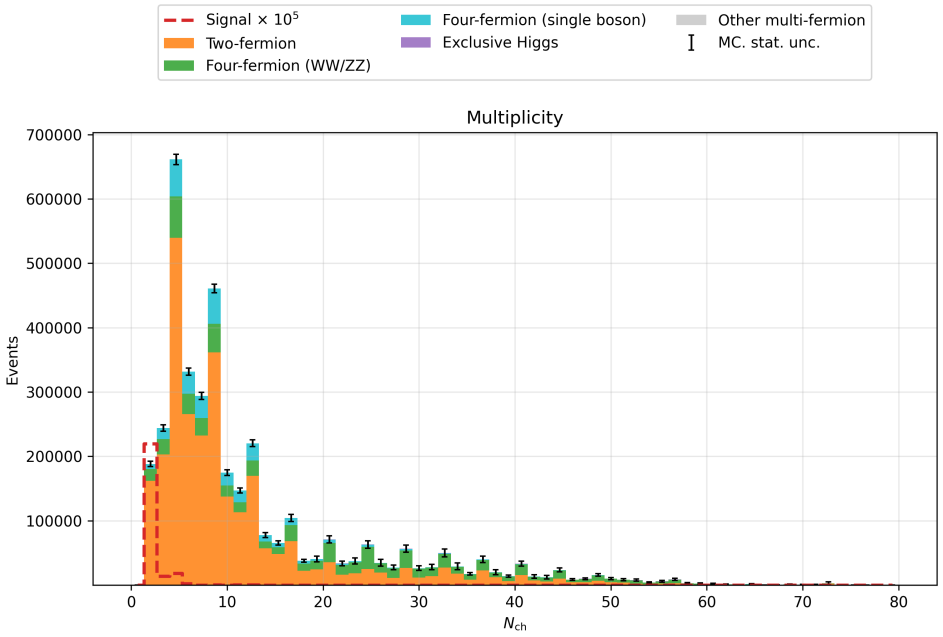


Figure 16 – Multiplicity distribution of the muon pair, N_{ch}, before event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^5 for visibility. The distribution is shown in the range $N_{\text{ch}} \in [0, 80]$. The signal manifests as a localized excess around $N_{\text{ch}} = 2$, while background processes populate a broader multiplicity spectrum. This separation motivates the application of a multiplicity cut to enhance the signal-to-background ratio by selecting events consistent with the initial $\mu^+\mu^-$ production. Error bars represent the Monte Carlo statistical uncertainties.

The TM requirement significantly suppresses the Four-fermion (single boson) background contribution as illustrated in Figure 30. In these processes, the final-state fermions are frequently produced with asymmetric momentum configurations or are strongly forward-peaked, leading to low TM for at least one of the muons.

In contrast, the signal topology involves the decay of a heavy boson into two muons with relatively balanced kinematics, resulting in both leptons typically exceeding the $p_T > 10\,\text{GeV}$ threshold. The requirement on the subleading muon TM is therefore particularly effective in rejecting background events with soft or collinear leptons.

The muon TM distributions before and after the application of the full event selection are presented in Figures 18, 19, 20 and 21, respectively.

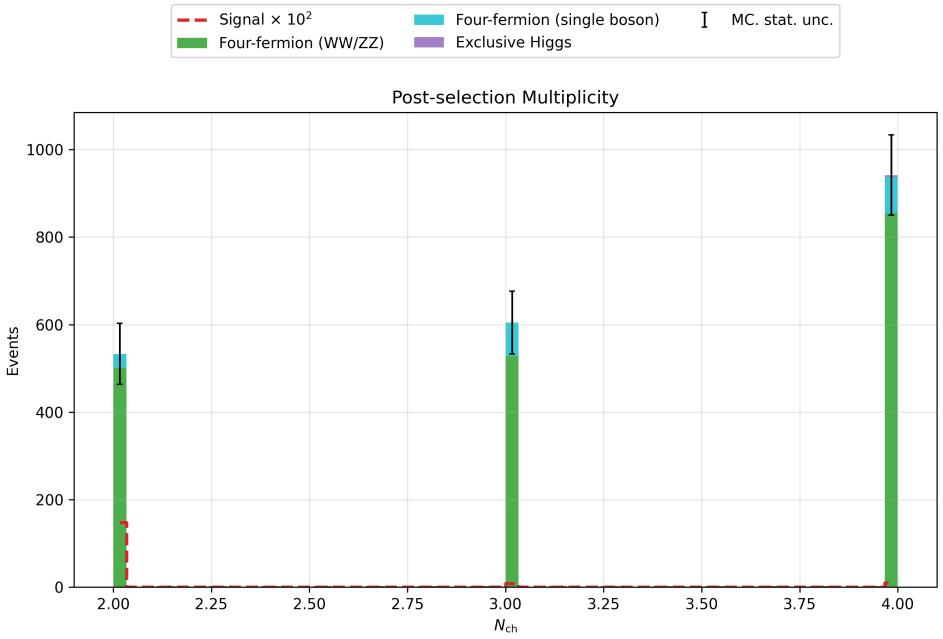


Figure 17 – Multiplicity distribution of the muon pair, N_{ch}, after event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^2 for visibility. The distribution is shown in the range $N_{\text{ch}} \in [2, 4]$. Error bars represent the Monte Carlo statistical uncertainties.

4.4.4 Muon Pair Energy

For two reconstructed muons with energies E_μ^{lead} and E_μ^{sublead}, the Muon Pair Energy is just the sum of the energy of the two muons:

$$E_{\mu\mu} = E_\mu^{\text{lead}} + E_\mu^{\text{sublead}} \tag{4.11}$$

Equivalently, in terms of four-momenta,

$$E_{\mu\mu} = (p_\mu^{\text{lead}} + p_\mu^{\text{lead}})^0 \tag{4.12}$$

where the superscript 0 denotes the energy component of the four-vector. The total energy of the muon pair corresponds to the energy of the reconstructed Z boson candidate. This quantity enters directly in the calculation of the recoil mass and is therefore a key observable in Higgs recoil analyses. Since muons are measured with excellent momentum resolution in lepton collider detectors, $E_{\mu\mu}$ is reconstructed with high precision.

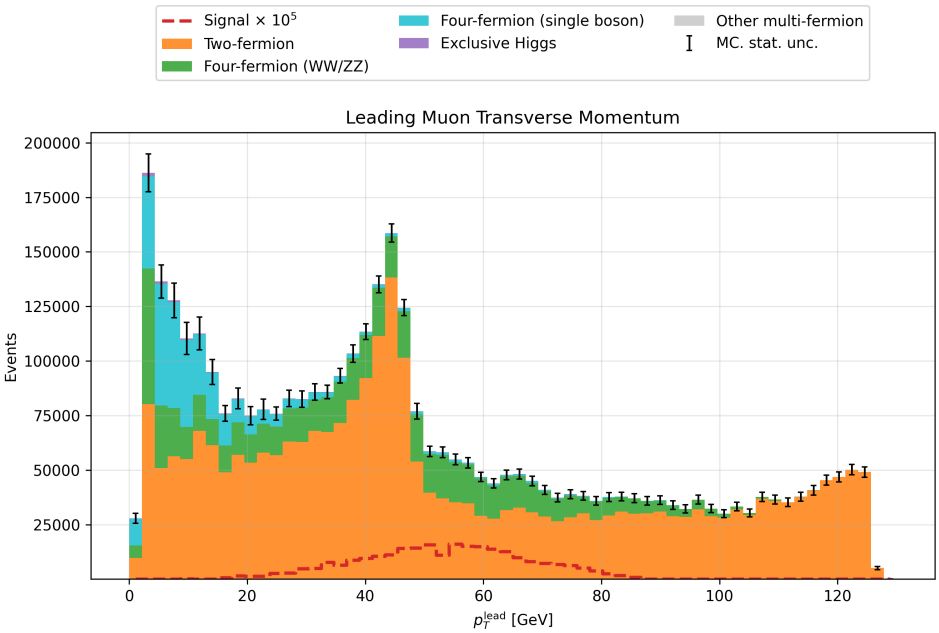


Figure 18 – Leading Muon Transverse Momentum distribution of the muon pair, p_T^{lead}, before event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^5 for visibility. The distribution is shown in the range $p_T^{\text{lead}} \in [0, 130]$ GeV. The signal manifests as a localized excess around $p_T^{\text{lead}} = 60$ and $p_T^{\text{sublead}} = 30$, while background processes populate a broader kinematic spectrum, specially lower values. This separation motivates the application of a individual muon TM cut to enhance the signal-to-background ratio by selecting muons with sufficient momentum. Error bars represent the Monte Carlo statistical uncertainties.

4.4.5 Missing Energy

The missing energy represents the fraction of the center-of-mass energy not carried by reconstructed visible particles. In the IHD channel, this quantity is sensitive to the presence of undetected final-state particles produced in the Higgs decay. While missing energy alone is affected by detector effects and initial-state radiation, it provides complementary information to the recoil mass in validating the invisible decay hypothesis. While the total muon-pair energy reflects the kinematics of the reconstructed Z boson, the missing energy quantifies the energy imbalance in the event and is sensitive to invisible final states. For the IHD, the missing energy is defined as:

$$E_{\text{miss}} = \sqrt{s} - E_{\mu\mu} \tag{4.13}$$

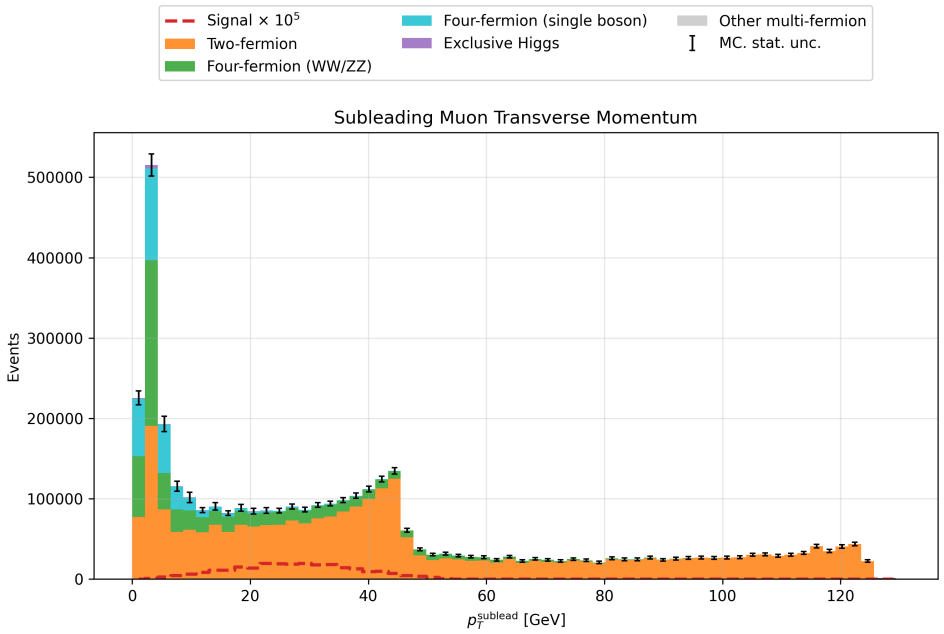


Figure 19 – Subleading Muon Transverse Momentum distribution of the muon pair, p_T^{sublead}, before event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^5 for visibility. The distribution is shown in the range $p_T^{\text{sublead}} \in [0, 130]$ GeV. Error bars represent the Monte Carlo statistical uncertainties.

this definition is only valid since the visible energy is equal to the muon pair energy $E_{\text{vis}} = E_{\mu\mu}$ because the only visible particles are the muon pair. For background events, other particles are produced in association with the muon pair $\mu^+\mu^-$ from either Z or h produced through the *Higgsstrahlung*. The `MCConsumer` algorithm was designed with adding the 4-momentum of any particle produced, if only two muons μ of opposite charge were detected, that the algorithm would calculate the $E_{\mu\mu}$, independent of their origin. Therefore, the missing energy for background events can be defined as:

$$E_{\text{miss}} = \sqrt{s} - E_{\mu\mu} - E_{\text{other}} \tag{4.14}$$

where $E_{\mu\mu}$ could be zero for non-muon-producing events and E_{other} is energy sum of all particles produced that are not the muon pair, so its values could zero or not depending on the amount of associated particles produced. The selection criteria chosen for the Missing Energy E_{miss} was for all values above 40 GeV:

$$E_{\text{miss}} > 40\,\text{GeV} \tag{4.15}$$

The missing energy requirement significantly reduces the Two-fermion and Four-fermion (single boson) background contributions, as illustrated in Figure 30. These processes

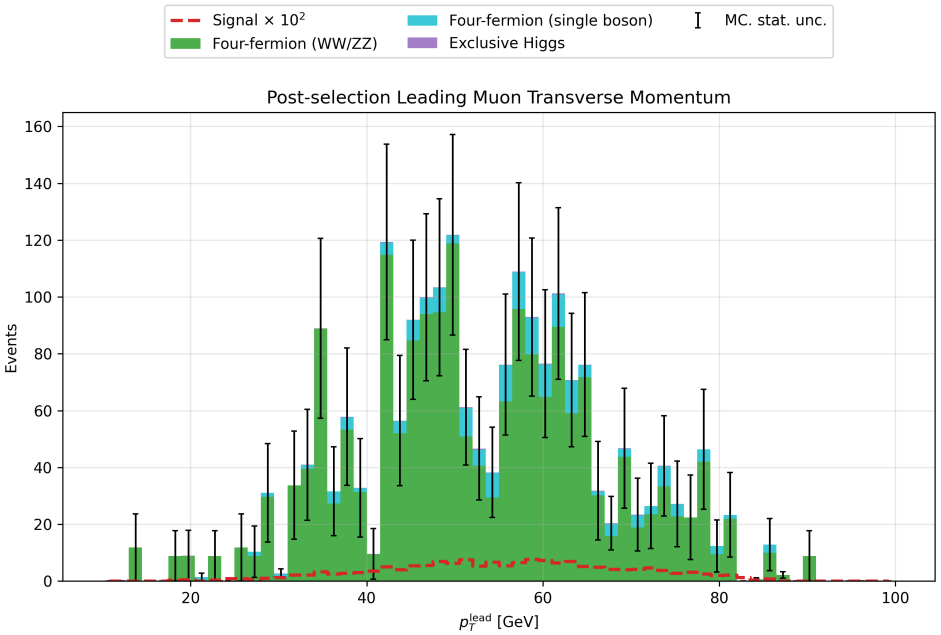


Figure 20 – Leading Muon Transverse Momentum distribution of the muon pair, p_T^{lead}, after event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^2 for visibility. The distribution is shown in the range $p_T^{\text{lead}} \in [10, 100]$ GeV. Error bars represent the Monte Carlo statistical uncertainties.

predominantly lead to fully visible final states, in which the total reconstructed energy closely matches the center-of-mass energy. Consequently, the missing energy in such events arises mainly from detector resolution effects or initial-state radiation escaping along the beam direction, and is typically small.

Requiring $E_{\text{miss}} > 40\,\text{GeV}$ therefore suppresses a large fraction of these background events. In contrast, the signal topology involves the production of invisible particles, resulting in a substantial genuine missing energy component. The chosen threshold thus enhances the signal-to-background separation by rejecting predominantly visible final states while retaining events consistent with the expected invisible recoil signature.

Figures 22 and 23 show the missing energy distribution before and after the full event selection, respectively.

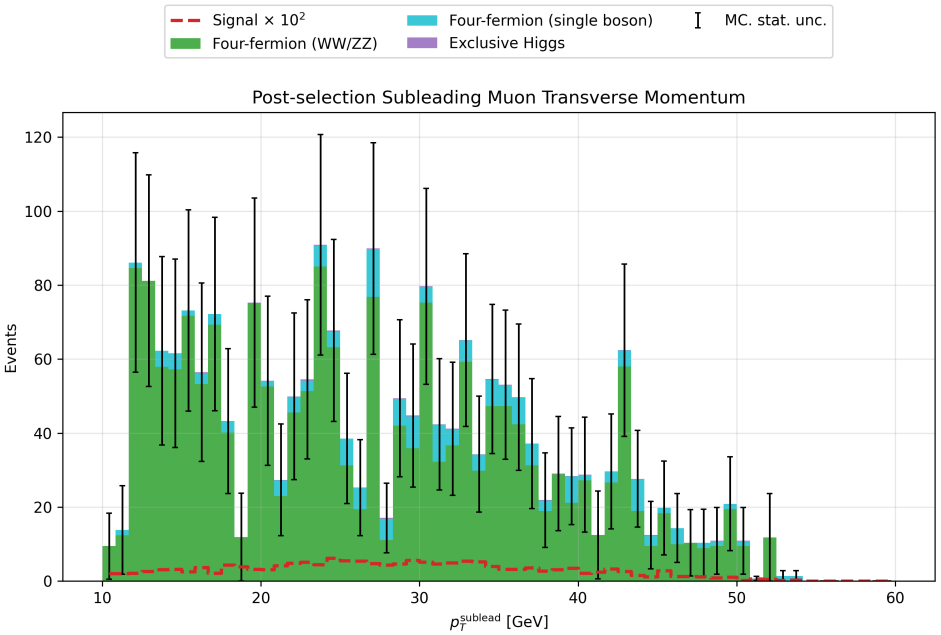


Figure 21 – Subleading Muon Transverse Momentum distribution of the muon pair, p_T^{sublead}, after event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^2 for visibility. The distribution is shown in the range $p_T^{\text{sublead}} \in [10, 60]$ GeV. Error bars represent the Monte Carlo statistical uncertainties.

4.4.6 Transverse Momentum

The TM of the dimuon system provides an important kinematic handle on the event topology. In the process $e^+e^- \to Zh$ with $Z \to \mu^+\mu^-$, the reconstructed Z boson recoils against the Higgs boson. As a consequence, the TM of the dimuon system reflects the global momentum balance of the event. An intermediary step before the invariant mass cut, is to cut the TM p_T of the dimuon system. First, a single-muon selection is imposed to ensuring that each reconstructed muon is sufficiently energetic. For a dimuon selection, a cut is applied to the TM of the dimuon system $p_{T\mu\mu}$, where, unlike the single-muon requirement, this cut acts on the magnitude of the vector sum of the two transverse momenta. Although each muon has positive TM by definition, their transverse components may partially cancel if they are emitted back-to-back. Consequently, $p_T^{\mu\mu}$ encodes the net transverse recoil of the reconstructed Z boson. Therefore, the transverse component of the dimuon momentum are given by

$$p_i^{\mu\mu} = p_i^{\text{lead}} + p_i^{\text{sublead}}, \qquad i \in \{x, y\}. \tag{4.16}$$

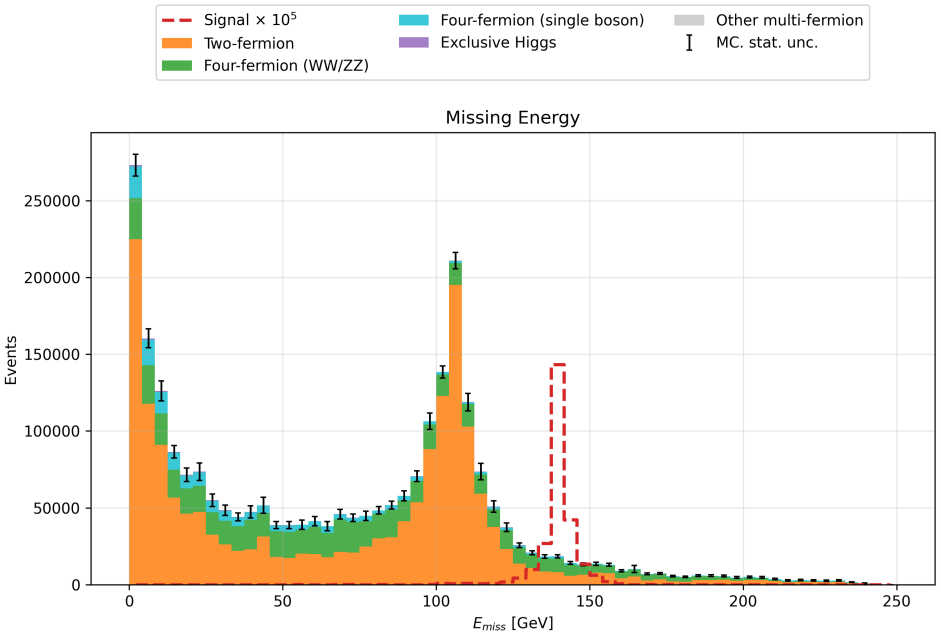


Figure 22 – Missing energy distribution of the muon pair, E_{miss}, before event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^5 for visibility. The distribution is shown in the range $E_{\text{miss}} \in [0, 250]$ GeV. Error bars represent the Monte Carlo statistical uncertainties.

then from equation 4.9, the momentum components are squared, added and then square-rooted, resulting in the TM of the dimuon system:

$$p_T^{\mu\mu} = \sqrt{(p_x^{\mu\mu})^2 + (p_y^{\mu\mu})^2} \tag{4.17}$$

the background processes such as cosmic-ray muons and $WW^* \to \mu^+\mu^-$ production can yield dimuon final states with kinematic properties distinct from the signal. To enhance sensitivity to the $e^+e^- \to Zh$ topology, a selection requirement is imposed on the dimuon TM:

$$20 \leq p_T^{\mu\mu} \leq 70\,\text{GeV} \tag{4.18}$$

The TM of the muon pair provides one of the strongest discriminating variables against the Two-fermion background, as illustrated in Figure 30. In the dominant process $e^+e^- \to Z/\gamma^* \to \mu^+\mu^-$, the final state consists of a two-body system with no recoil against an additional massive particle. Consequently, the muons are produced predominantly back-to-back in the transverse plane, resulting in a small net TM of the dimuon system. Non-zero values arise mainly from initial-state radiation and detector effects.

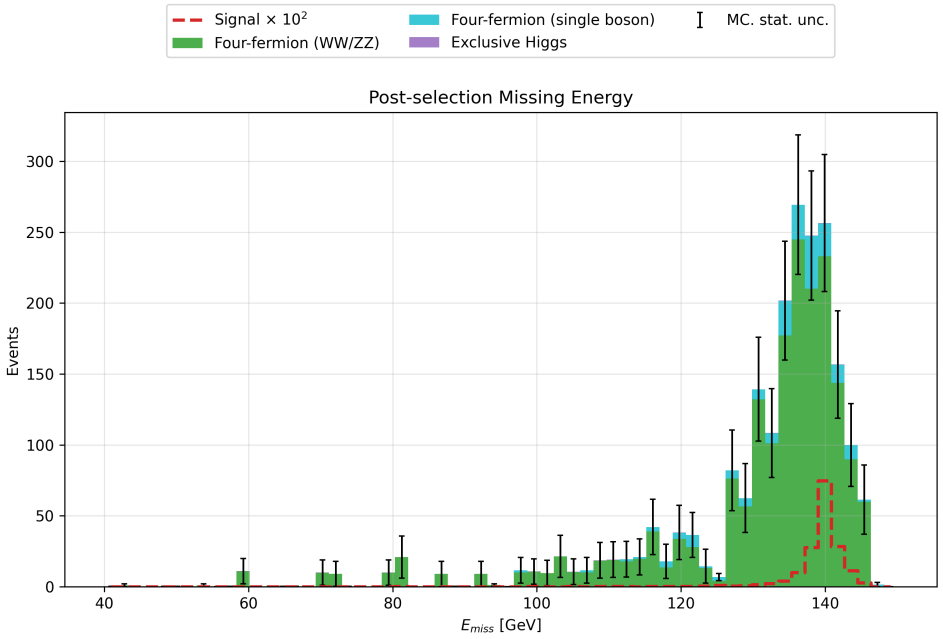


Figure 23 – Missing energy distribution of the muon pair, E_{miss}, after event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^2 for visibility. The distribution is shown in the range $E_{\text{miss}} \in [40, 150]\,\text{GeV}$. The signal manifests as a localized excess around $E_{\text{miss}} = 140\,\text{GeV}$, while background processes populate a lower missing energy spectrum. This separation motivates the application of a missing energy cut to enhance the signal-to-background ratio by selecting events consistent with sufficient missing energy. Error bars represent the Monte Carlo statistical uncertainties.

In contrast, for the signal process $e^+e^- \to Zh$ with $h \to \text{inv}$ at $\sqrt{s} = 250\,\text{GeV}$, the Z boson recoils against a 125 GeV Higgs boson. This recoil imparts a substantial TM to the dimuon system, populating an intermediate $p_T^{\mu\mu}$ region characteristic of the signal topology.

The requirement $20 \leq p_T^{\mu\mu} \leq 70\,\text{GeV}$ therefore efficiently suppresses the large concentration of low-p_T Two-fermion events while retaining the kinematic region expected for Zh production. This explains the observed reduction of the Two-fermion contribution to approximately 2% of its previous value.

Figures 24 and 25 shows the dimuon TM distribution before and after the full event selection, respectively.

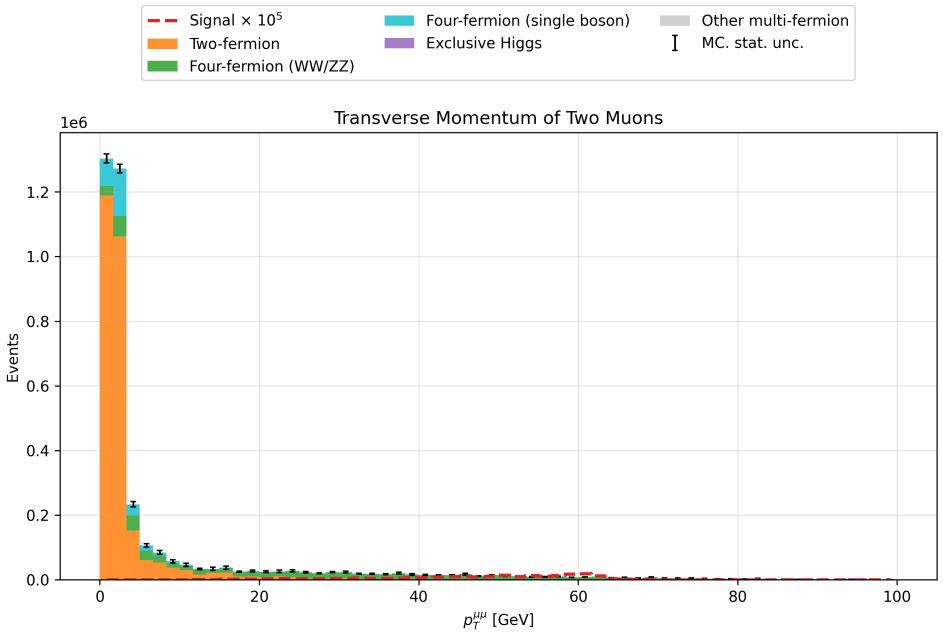


Figure 24 – Dimuon TM distribution of the muon pair, $p_T^{\mu\mu}$, before event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^5 for visibility. The distribution is shown in the range $p_T^{\mu\mu} \in [0, 100]$ GeV. The signal manifests as a broad presence around $20 \leq p_T^{\mu\mu} \leq 70$ GeV, while background processes populate as a localized excess at a lower kinematic spectrum. This separation motivates the application of a dimuon TM cut to enhance the signal-to-background ratio by selecting events consistent with sufficiently dimuon momentum. Error bars represent the Monte Carlo statistical uncertainties.

4.4.7 Reconstructed Total Momentum

The magnitude of the three-momentum of the dimuon system, $\vec{p}_{\mu\mu}$, is required for the reconstruction of kinematic observables such as the invariant mass and the recoil mass. It is obtained from the combined spatial components of the dimuon momentum defined in equation 4.16. The calculation of the total magnitude of the 3-momentum of the Muon pair $\vec{p}_{\mu\mu}$ is:

$$|\vec{p}_{\mu\mu}| = \sqrt{(p_x^{\mu\mu})^2 + (p_y^{\mu\mu})^2 + (p_z^{\mu\mu})^2}. \tag{4.19}$$

where the momentum squared of each dimensional comes from the definition from equation 4.16. The 3-momentum fully characterizes the spatial momentum of the reconstructed $Z \to \mu^+\mu^-$ candidate and enters directly in the computation of the invariant mass $m_{\mu\mu}$ as well as the recoil mass m_{recoil} defined in the following sections.

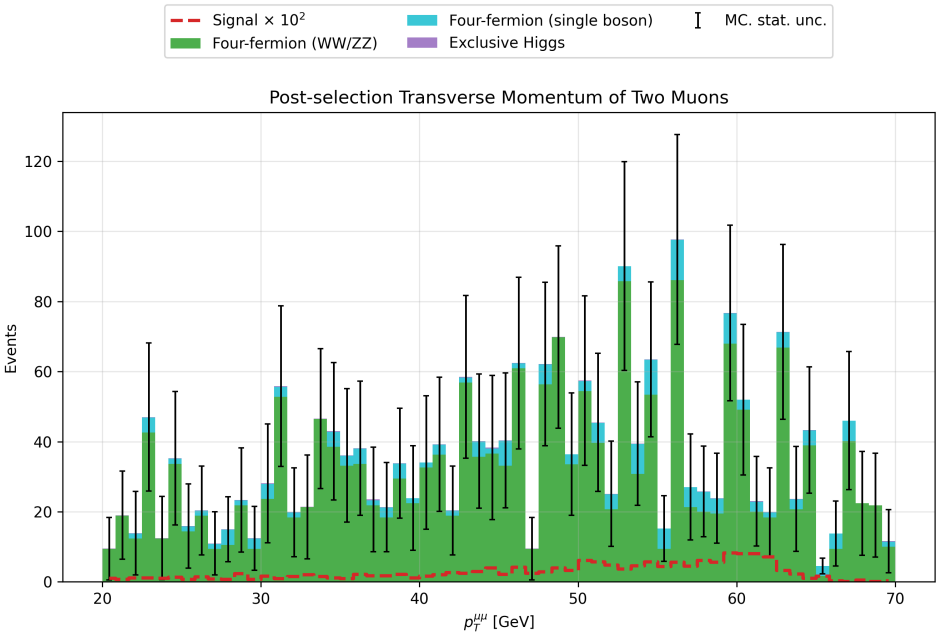


Figure 25 – Dimuon TM distribution of the muon pair, $p_T^{\mu\mu}$, after event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^2 for visibility. The distribution is shown in the range $p_T^{\mu\mu} \in [20, 70]$ GeV. Error bars represent the Monte Carlo statistical uncertainties.

4.4.8 Invariant Mass

The squared Transverse Mass m_T of a single particle is found by the following relation[2]:

$$m^2 = E^2 - |\vec{p}|^2 \tag{4.20}$$

this is also invariant to boosts along the proton beam, so it can also be called squared invariant mass. Since the expected system only produces the pair of muons for the visible system the invariant mass $m_{\mu\mu}$ is equivalent to the visible mass, therefore:

$$m_{\mu\mu}^2 = E_{\mu\mu}^2 - |\vec{p}_{\mu\mu}|^2 \tag{4.21}$$

square-rooting this equation results in the desired invariant mass $m_{\mu\mu}$ formula:

$$m_{\mu\mu} = \sqrt{E_{\mu\mu}^2 - |\vec{p}_{\mu\mu}|^2} \tag{4.22}$$

given that the production of the pair of muons comes from the decay of Z boson, the reconstructed invariant mass should be the same values as the mass of the Z boson m_Z.

[2] Values for c and $\hbar$ has been set to 1.

For background events, more particles can be produced alongside the muon pair, breaking the relation $m_{\text{vis}} \neq m_{\mu\mu}$. Previous studies are divide into two similar but not equal choices for the selection criteria. First, some studies chose to limit the $m_{\mu\mu}$ into a interval with the m_Z roughly at its center, e.g. $m_{\mu\mu} \in [75, 105]\,\text{GeV}$, with varied choices for the size of the interval (2). For the second choice, the one used in this study, was the selection criteria based on the absolute difference between the invariant mass the Z boson mass (3):

$$|m_{\mu\mu} - m_Z| < \Delta m_Z \tag{4.23}$$

where $m_Z = 91.1880\,\text{GeV}$ and letting the interval of $\Delta m_Z = 10\,\text{GeV}$. The invariant mass requirement selects dimuon pairs consistent with an on-shell Z boson decay. For the signal process $e^+e^- \to Zh$ with $h \to \text{inv}$, the muons originate from $Z \to \mu^+\mu^-$, and therefore the invariant mass is expected to peak sharply at m_Z.

In contrast, the Two-fermion background includes contributions from off-shell γ^* exchange and Z/γ^* interference, leading to a broader invariant mass distribution. The requirement $|m_{\mu\mu} - m_Z| < 10\,\text{GeV}$ removes a substantial fraction of these non-resonant events.

Similarly, in Four-fermion (WW/ZZ) processes, only a subset of events contains a genuine $Z \to \mu^+\mu^-$ decay. In particular, WW production does not involve a Z resonance in the dimuon system, resulting in a broad invariant mass spectrum. The mass window therefore significantly reduces these backgrounds, explaining the observed reduction to approximately 10% of their previous yield.

Figures 26 and 27 respectively shows the invariant mass before and after the all the events selection.

4.4.9 Recoil Mass

The ILC is an e^+e^- collider, the recoil system is defined as the difference between the energy of the reconstructed $Z \to \mu^+\mu^-$ system and the initial energy of the system $\sqrt{s}$, the goal is to check whenever the Z boson originated from the Higgs, for a collision energy of $\sqrt{s} = 250\,\text{GeV}$:

$$p_{\text{recoil}} = p_{ee} - p_{\mu\mu} \tag{4.24}$$

where $p_{e^+e^-}$ is the initial four-momentum, the squared recoil mass is defined as the squared invariant mass of this four-vector:

$$m^2_{\text{recoil}} = p^2_{\text{recoil}} = (p_{ee} - p_{\mu\mu})^2 \tag{4.25}$$

$$= p^2_{ee} + p^2_{\mu\mu} - 2p_{ee} \cdot p_{\mu\mu} \tag{4.26}$$

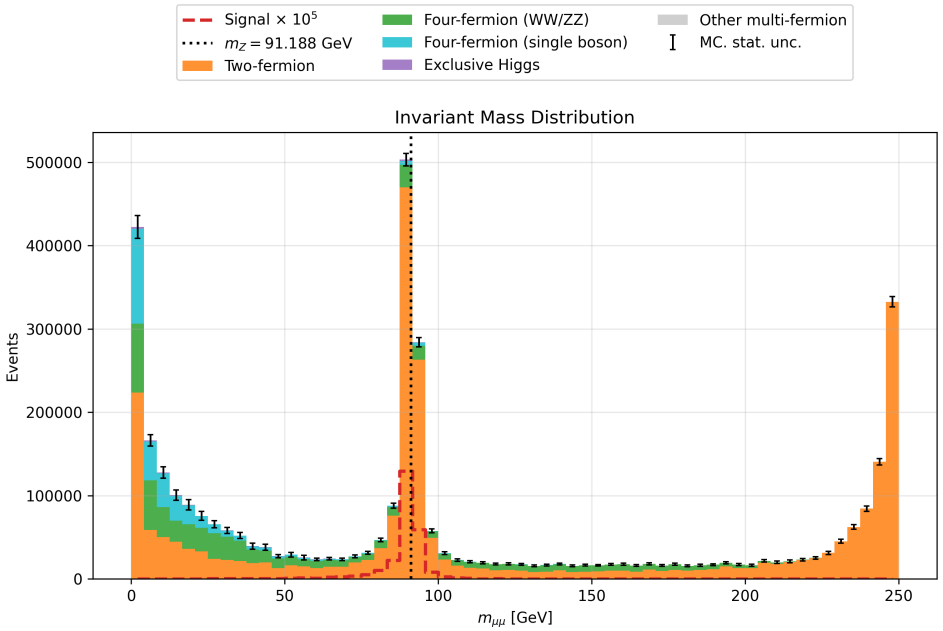


Figure 26 – Invariant mass distribution of the muon pair, $m_{\mu\mu}$, before event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^5 for visibility. The distribution is shown in the range $m_{\mu\mu} \in [0, 250]\,\text{GeV}$. The black vertical dashed line indicates the Z boson mass, $m_Z = 91.188\,\text{GeV}$. The signal manifests as a localized excess around m_Z, while background processes populate a broader kinematic spectrum. This separation motivates the application of a invariant mass window to enhance the signal-to-background ratio by selecting events consistent with the $Z \to \mu^+\mu^-$ production hypothesis. Error bars represent the Monte Carlo statistical uncertainties.

now get the total 4-Momentum of the muon system $p_{\mu\mu} = p_{\mu^+} + p_{\mu^-}$ and by definition, the square of this four-vector is the invariant mass squared of the muon pair $p^2_{\mu\mu} = m^2_{\mu\mu}$. Subtract with the momentum of the system to the recoil momentum, however the initial three-momentum of the system is zero $\vec{p}_{e^+e^-} = 0$ and this leaves the scalar product $p_{e^+e^-} \cdot p_{\mu^+\mu^-}$ as:

$$p_{ee} \cdot p_{\mu\mu} = E_{ee}E_{\mu\mu} - \vec{p}_{ee} \cdot \vec{p}_{\mu\mu} \tag{4.27}$$

$$= \sqrt{s}E_{\mu\mu} \tag{4.28}$$

only to calculate the difference between energies:

$$p_{ee} = (\sqrt{s}, 0, 0, 0) \tag{4.29}$$

$$p^2_{ee} = (\sqrt{s})^2 = s \tag{4.30}$$

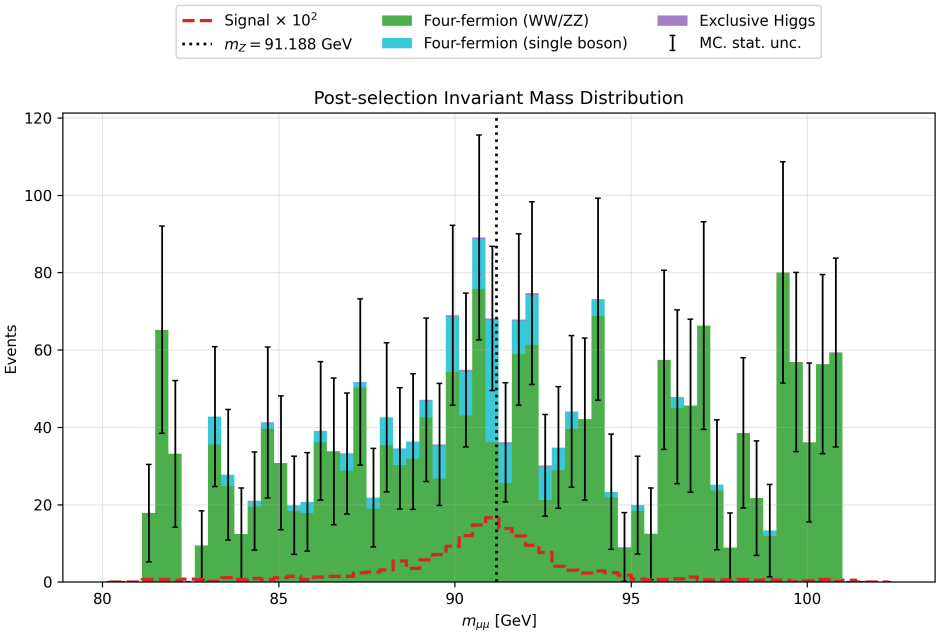


Figure 27 – Invariant mass distribution of the muon pair, $m_{\mu\mu}$, after event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^2 for visibility. The distribution is shown in the range $m_{\mu\mu} \in [80, 102.5]$ GeV. The black vertical dashed line indicates the Z boson mass, $m_Z = 91.188$ GeV. Error bars represent the Monte Carlo statistical uncertainties.

Therefore the squared recoil-mass definition (2) is:

$$m^2_{\text{recoil}} = (\sqrt{s} - E_{\mu\mu})^2 - |\vec{p}_{\mu\mu}|^2 \tag{4.31}$$

$$= s + m^2_{\mu\mu} - 2\sqrt{s}E_{\mu\mu} \tag{4.32}$$

where $\sqrt{s}$ is the collider energy, $m_{\mu\mu}$ the reconstructed Z boson mass and $E_{\mu\mu}$ the energy of the visible Z decay. Square-rooting the equation 4.31, result in the recoil mass formula that was used for analysis and plotting:

$$m_{\text{recoil}} = \sqrt{s + m^2_{\mu\mu} - 2\sqrt{s}E_{\mu\mu}} \tag{4.33}$$

the result is the a histogram where the the most common event also coincides with the Higgs mass $m_h = 125.20$ GeV in graph 28. Similarly to the invariant mass 4.4.8, the selection criteria for the recoil mass m_{recoil} is done with a fixed interval with the Higgs mass roughly at its center (2, 5, 4). Following the same logic as the invariant mass, the selection criteria for this work was determined based on the absolute difference between

the calculated value and expected value:

$$|m_{\text{recoil}} - m_h| < \Delta m_h \tag{4.34}$$

where $\Delta m_h = 10\,\text{GeV}$. The intrinsic Higgs boson width ($\Gamma_H \sim 4\,\text{MeV}$) is negligible compared to the experimental recoil mass resolution. The observed width of the m_{recoil} distribution is therefore dominated by detector resolution, beam energy spread, and initial-state radiation effects.

The chosen window corresponds approximately to a $\mathcal{O}(1\text{–}2\sigma)$ interval around the expected Higgs mass, balancing signal efficiency against background rejection. A narrower window would reduce signal yield, while a wider window would admit additional background contributions from non-resonant processes.

Figures 28 and 29 respectively shows the multiplicity before and after the all the events selection.

The recoil mass requirement provides the most powerful discrimination against the remaining Two-fermion background. In the signal process $e^+e^- \to Zh$ with $h \to \text{inv}$, the Higgs boson is produced in association with an on-shell Z boson in a two-body final state. As a result, the invariant mass of the system recoiling against the reconstructed dimuon pair peaks at $m_H \approx 125\,\text{GeV}$.

In contrast, for the dominant Two-fermion background $e^+e^- \to Z/\gamma^* \to \mu^+\mu^-$, no additional heavy particle is produced. The muon pair therefore carries nearly the full center-of-mass energy, and the recoil mass is close to zero, apart from effects due to initial-state radiation and detector resolution. Consequently, these events do not populate the region around m_H.

The requirement $|m_{\text{recoil}} - m_H| < 10\,\text{GeV}$ thus effectively selects events consistent with Zh production and eliminates the remaining Two-fermion contribution.

4.4.10 Global Cutflow Performance

Having examined the impact of each selection requirement individually, the overall effect of the full selection sequence is summarized in Figure 30, that has the yield values for all signal background categories and Figure 31, presents the relative evolution between cuts. The relative cutflow evolution was created to measure the suppression of yields between each cut defined by the formula:

$$N_i/N_{i-1} \tag{4.35}$$

where N_i denotes the weighted yield after the i-th selection requirement. This quantity isolates the incremental suppression power of each cut, independently of the previous normalization.

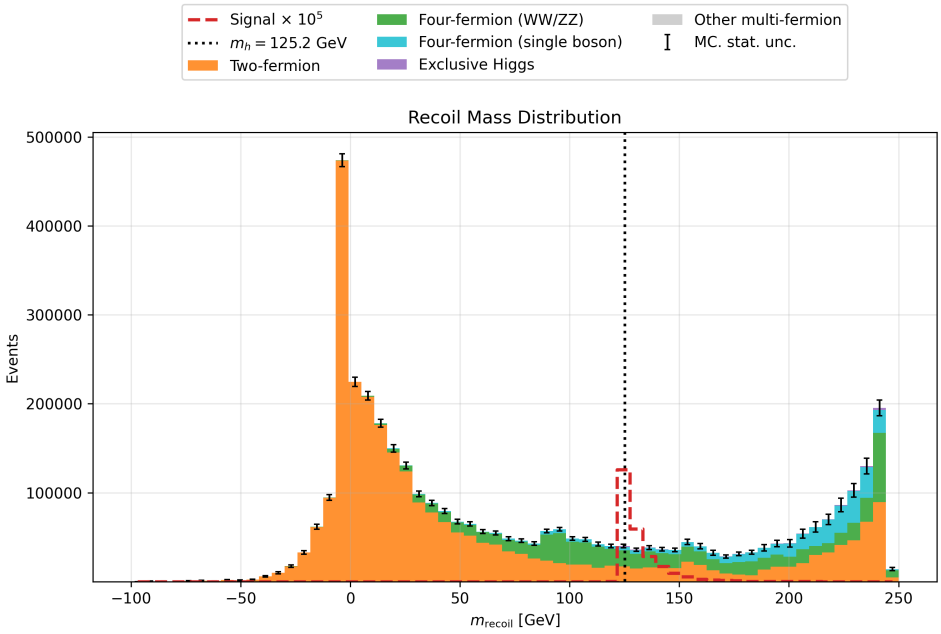


Figure 28 – Recoil Mass distribution of the muon pair, m_{recoil}, before event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^5 for visibility. The distribution is shown in the range $m_{\text{recoil}} \in [-100, 250]$ GeV. The black vertical dashed line indicates the h boson mass, $m_h = 125.20$ GeV. The signal manifests as a localized excess around m_h, while background processes populate a broader kinematic spectrum. This separation motivates the application of a recoil mass window to enhance the signal-to-background ratio by selecting events consistent with the two-body Zh production hypothesis. Error bars represent the Monte Carlo statistical uncertainties.

Figure 30 shows that the dominant backgrounds at the preselection level are the two and four-fermion processes, particularly the two-fermion processes. As the selection progresses, a strong hierarchical suppression is observed. Multi-fermions processes are completely suppressed at early stages in the multiplicity cut. The Two-fermion processes remains present in most stages, until it's completely suppress in the recoil mass cut, while four-fermion backgrounds persist longer due to their genuine dimuon topologies and missing energy components.

The relative cutflow evolution in Figure 31 highlights the discriminating power of each selection step. Signal efficiencies remain comparatively stable across most cuts, while specific background categories exhibit sharp reductions, indicating that the selection strategy is well aligned with the underlying signal topology.

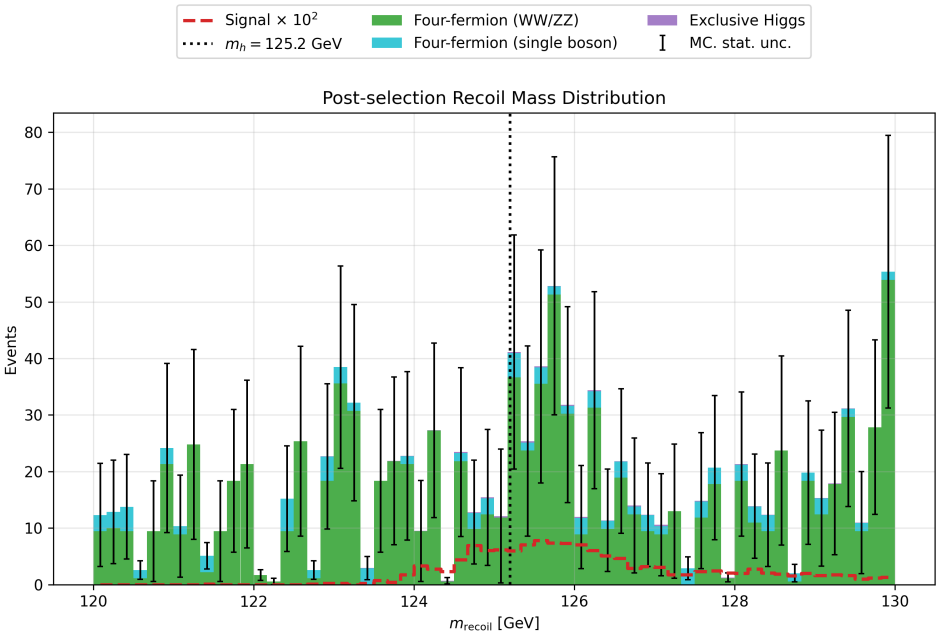


Figure 29 – Recoil Mass distribution of the muon pair, m_{recoil}, after event selection. Stacked histograms show the expected background contributions, color-coded according to the process categories in Table 4. The invisible Higgs signal (red dashed line) is scaled by a factor of 10^2 for visibility. The distribution is shown in the range $m_{\text{recoil}} \in [120, 130]\,\text{GeV}$. The black vertical dashed line indicates the h boson mass, $m_h = 125.20\,\text{GeV}$. Error bars represent the Monte Carlo statistical uncertainties.

Lastly, a comparison between Figures 14 and 32, illustrates the effectiveness of the employed cuts. Starting with a contribution of $6.77 \times 10^{-5}\%$ to total yield, the signal ended up contributing with $8.00 \times 10^{-2}\%$, a difference in three orders of magnitude.

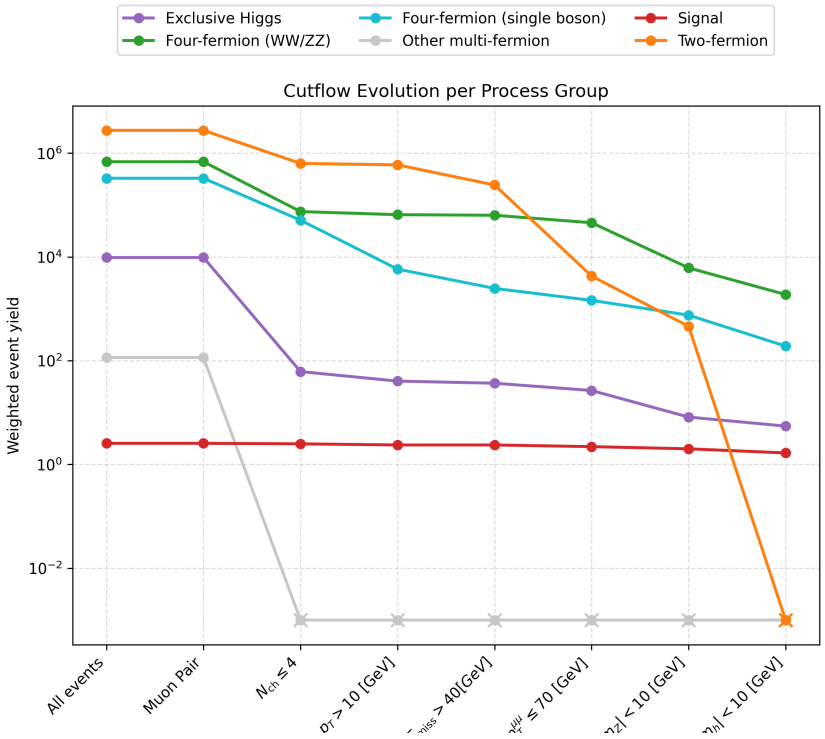


Figure 30 – Cutflow evolution of weighted event yields for each physics process group under the sequential application of the selection requirements. The logarithmic scale highlights the progressive suppression of dominant background contributions. The "x" markers indicates that the contribution for an specific category has been completely suppressed.

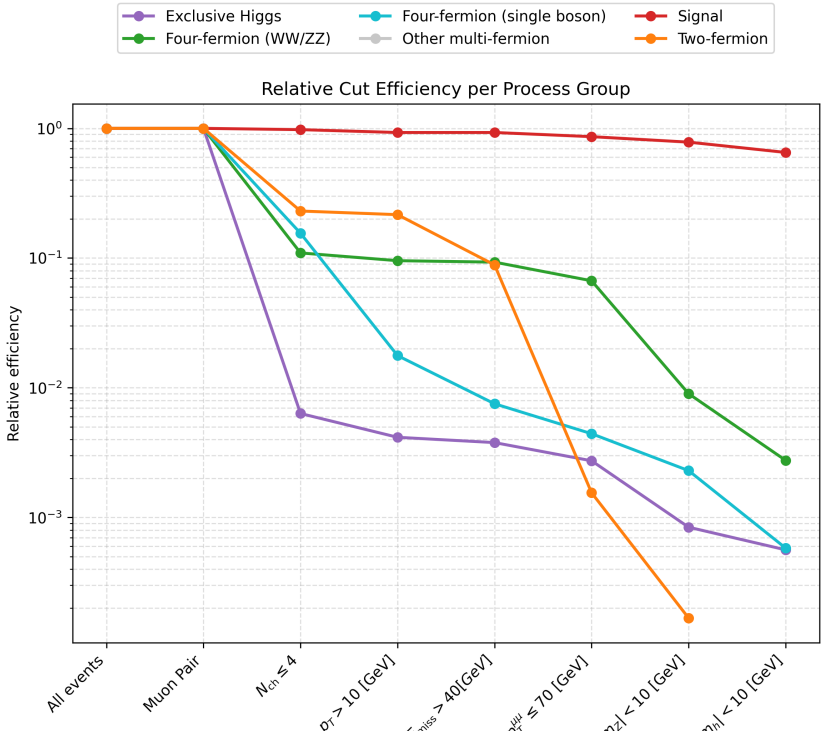


Figure 31 – Relative cut-by-cut survival efficiency N_i/N_{i-1} for each process group. The logarithmic scale emphasizes the discriminating power of individual selection requirements across signal and background categories.

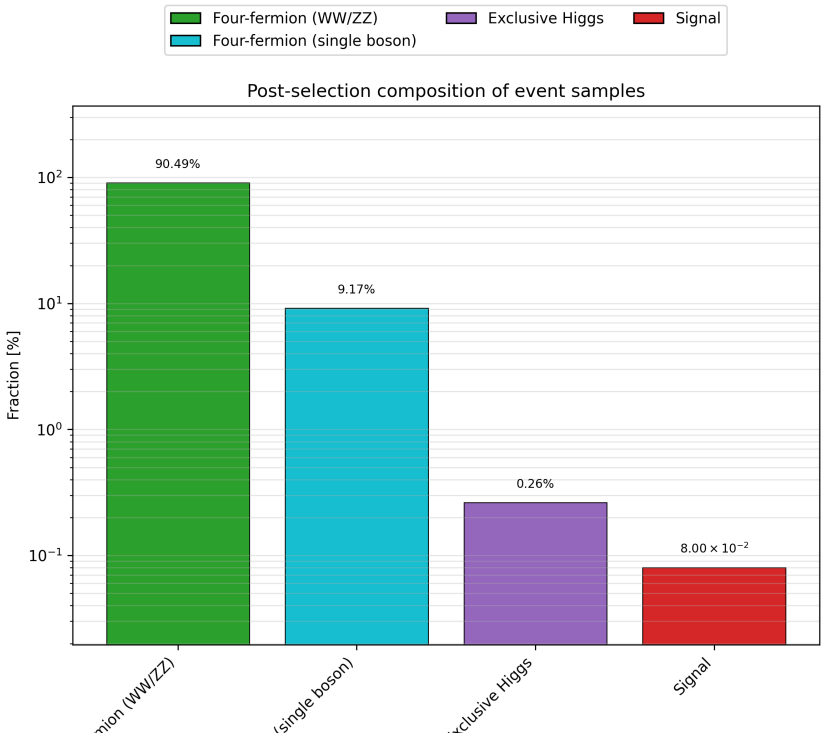


Figure 32 – Percentage composition of the simulated events in log after all the cuts from Tables 17 and 18, each bar is a category as described in Table 4.

5 Results

This chapter presents the expected sensitivity to the IHD channel in the process $e^+e^- \to Zh$, with $Z \to \mu^+\mu^-$ at $\sqrt{s} = 250\,\text{GeV}$. All results are obtained using fully simulated Monte Carlo samples within the workflow of the Key4hep ecosystem. The following limitations are explicitly acknowledged in order to clearly define the scope of this work:

- Background yields are normalized to the assumed luminosity,
- No control-region normalization is performed,
- No MVA or cut optimization scan was performed,
- No control regions or data-driven constraints were implemented (MC-only analysis),
- No full detector systematic uncertainties evaluation was carried out.
- No complete statistical model with multiple bins or shape fit was constructed,
- Not a full profile-likelihood or CL exclusion limit analysis,
- No comparison across different generators or tunes was performed.

The two key fundamental aspect of the IHD analysis, were the invariant mass and the recoil mass technique technique, that allows the reconstruction of the Z and Higgs boson respectively, both of them are independently of its decay products. The combination of these two techniques is fundamental in connecting the decay chain of these two processes, with the goal of trying to stablish simplified projected sensitivity estimate $\text{BR}(h \to \text{inv})$, the invariant mass technique constrains the decay chain to $Z \to \mu^+\mu^-$ and the recoil mass technique constrains to $h \to \text{inv}$, while the other selection criteria served the purpose of reducing background samples by establishing minimum criteria for the visible and invisible parts of the system, such as ISR, extra particles, minimum momenta and energy for the each muon individually and as a pair. Signal events cluster around the Z and Higgs boson mass, while background processes exhibit broader kinematic distributions.

The cuts used to build the cutflow Tables 17 and 18 where inspired by lepton channel analysis performed at (2, 3, 4, 5, 6). The tables shows that the selected cuts are enough reduce the background yield by a factor of a 10^5, the signal yield still remained with 63.2% of its originals samples. This results in a substantial improvement in the signal-to-background ratio, confirming the robustness of the cut-based strategy

Background samples included all decays chains where the only muons presents were a muon pair of opposite charge $\mu^+\mu^-$, besides that, any or no particle, whenever visible or not, could be present in the system. After the full event selection, Figure 32 shows the final sample composition is dominated by two processes, first by four-fermion (WW/ZZ) with 90.5% total yield and second four-fermion (single boson) 9.2% total yield. This behavior is expected from the underlying electroweak production mechanisms at $\sqrt{s} = 250\,\text{GeV}$.

The four-fermion (WW/ZZ) contribution, particularly from $e^+e^- \to ZZ \to \mu^+\mu^-\nu\bar{\nu}$, constitutes an irreducible background, as it reproduces exactly the same visible final state topology as the signal. The separation between this background and the signal relies entirely on the recoil mass distribution, making it the dominant physical background after selection.

All other processes are efficiently suppressed by the combination of charged multiplicity, missing energy, TM, and mass window requirements.

5.0.0.1 Validation and Sanity Checks

The resulting value is the mass of the Z boson $m_Z = 91.1880\,\text{GeV}$, as the graph 26 shows that the most common value for the Signal invariant mass coincides with m_Z, Figure 32 shows that the "Other multi-fermion" and "Two-fermion" categories has been completely removed after the selection process.

Negative values of the reconstructed recoil mass arise when detector resolution effects, ISR, or hard neutrino emission cause the reconstructed dimuon four-momentum to violate exact energy–momentum conservation, Figure 28 does not shows that rarely, some background samples had values that goes beyond the expected 3-momentum for $\sqrt{s} = 250\,\text{GeV}$, inevitably resulting in a small fraction of values that result in negative values for the reconstruction of the recoil mass $m_{\text{recoil}} < 0\,\text{GeV}$, this is a consequence of resolution and ISR effects that perturb the reconstructed kinematics that didn't affect the samples analysis since the events to fall outside the signal region. These effects predominantly affect two and four-fermion background processes, a further investigation is needed to find out what final states are present in the negative recoil mass interval. In contrast, signal events remain sharply peaked around the Higgs boson mass, confirming the correct implementation of the recoil mass reconstruction.

5.1 Cuts Evolution Metrics

To evaluate the evolution of the signal and background samples, the evolution metrics of efficiency ϵ, rejection R, signal-to-background ratio S/B and FoM are introduced

Table 19 – Difference between steps in efficiency ϵ and ratio between steps rejection R_{step}, fraction between old and new values and FoM_{step}, fraction between new and old values. The initial values are

Requirement	$\Delta\epsilon$	R_{step}	S/B_{step}	FoM_{step}
All Events	0.000	1.000	1.000	1.000
Muon Pair	0.031	30.739	29.799	5.375
N_{ch}	0.022	4.963	4.852	2.178
Muon $p_T > 10$	0.047	1.142	1.085	1.016
E_{miss}	0.000	2.150	2.150	1.466
$20 \leq p_T^{\mu\mu} \leq 70\,\text{GeV}$	0.064	6.023	5.595	2.280
$\lvert m_{\mu\mu} - m_Z \rvert < 10\,\text{GeV}$	0.077	6.970	6.324	2.395
$\lvert m_{\text{recoil}} - m_h \rvert < 10\,\text{GeV}$	0.127	3.546	2.953	1.568
Total (cumulative)	0.368	55764.7	35261.6	149.2

to demonstrate how each cut contributed to the cutflow table:

$$\Delta\epsilon = \epsilon_{i-1} - \epsilon_i \tag{5.1}$$

values for $\Delta\epsilon$ represent the percentage of lost signal samples. For the rejection ratio:

$$R_{\text{step}} = \frac{R_{i-1}}{R_i} \tag{5.2}$$

the rejection increases with each step, therefore the evolution should measure the ratio between old and new values. The signal-to-background ratio and FoM increase with each step, then:

$$S/B_{\text{step}} = \frac{(S/B)_i}{(S/B)_{i-1}} \tag{5.3}$$

$$\text{FoM}_{\text{step}} = \frac{\text{FoM}_i}{\text{FoM}_{i-1}} \tag{5.4}$$

where i is the current cut and $i-1$ is the previous cut, Table 19 compiles the metrics evolution with each step. It is important to note that the first cut, the condition that the only muons present are a dimuon pair of opposite charge $\mu^+\mu^-$, had the highest vales for the rejection $R_{\text{step}} \approx 30.7$ and $\text{FoM}_{\text{step}} = 5.3$ steps as seen in Figure 33, this is due the fact this selection is in the simulation process and this condition is what determines what data produced should or should not be saved into the `ROOT` for further analysis. Therefore, the muon pair step is comparing before and after any selection criteria is done, resulting in great reduction of the background samples.

The recoil mass cut had the highest efficiency loss, with a $\Delta\epsilon = 0.127$ means a loss of 12.7% of in relation to the previous yield, while The invariant mass cut had the highest background reduction between steps $R_{\text{step}} \approx 7.0$ and connecting to the previous Muon Pair TM $p_T^{\mu\mu}$ cut led to $R_{\text{step}} \approx 6.0$, also worth noting that they also posses the two highest $\text{FoM}_{\text{step}} \approx 2.4, 2.3$ respectively, both cuts not only constraint the muons to

the decay state $Z \to \mu^+\mu^-$ but also makes sure the muons present have enough TM to reconstruct the Z boson.

The last row in Table 19 is the cumulative evolution of all metrics, the total rejection ratio means that the initial background samples are bigger by four orders of magnitude, $R_0 \lesssim 10^4 R_f$. The signal-to-background ratio also increased by more than four orders of magnitude $(S/B)_f \gtrsim 10^4 \times (S/B)_0$ as seen in Figure 33, lastly, the FoM increased by an order of $\text{FoM}_f \approx 150 \times \text{FoM}_0$.

A further analysis into Table 33 shows a correlation between the rejection and signal-to-background ratio evolution, equation 33 can be expanded into:

$$S/B_{\text{step}} = \frac{(S/B)_i}{(S/B)_{i-1}} = \frac{S_i}{S_{i-1}} \frac{B_{i-1}}{B_i} \tag{5.5}$$

Since the analysis is performed in the regime $S \ll B$, the optimization strategy aims to preserve signal while suppressing background. Consequently, the signal step efficiency remains close to unity and it's easier to quantify its evolution by the difference between efficiency than the ratio, specially because the signal order of magnitude doesn't change, resulting in the relation the $S_i/S_{i-1} \propto \Delta\epsilon$. The dominant contribution to the purity improvement arises from background rejection. Therefore,

$$(S/B)_{\text{step}} \approx R_{\text{step}} \tag{5.6}$$

showing the correlation between the rejection and signal-to-background ratio evolution.

5.2 Illustrative Uncertainty Propagation on BR Extraction

As a proof-of-principle, this section dedicated to performing a simplified uncertainty propagation to illustrate the relative impact of statistical and normalization uncertainties on the BR extraction, limiting the uncertainties to three variables:

- σ_{signal}: MC finite statistics, a statistic derived uncertainty
- σ_{Zh}: The *Higgstrahlung* cross-section, a parametric uncertainty
- $\text{BR}(Z \to \mu^+\mu^-)$: The BR for $Z \to \mu^+\mu^-$, a parametric uncertainty

no systematic uncertainties present at Table 20 were considered in this analysis. To extract the BR for Higgs to invisible, $\text{BR}(h \to \text{inv})$, from equation 3.10, rewriting the signal cross-section leads to:

$$\text{BR}(h \to \text{inv}) = \frac{\sigma_{\text{signal}}}{\sigma_{Zh} \times \text{BR}(Z \to \mu^+\mu^-)} \tag{5.7}$$

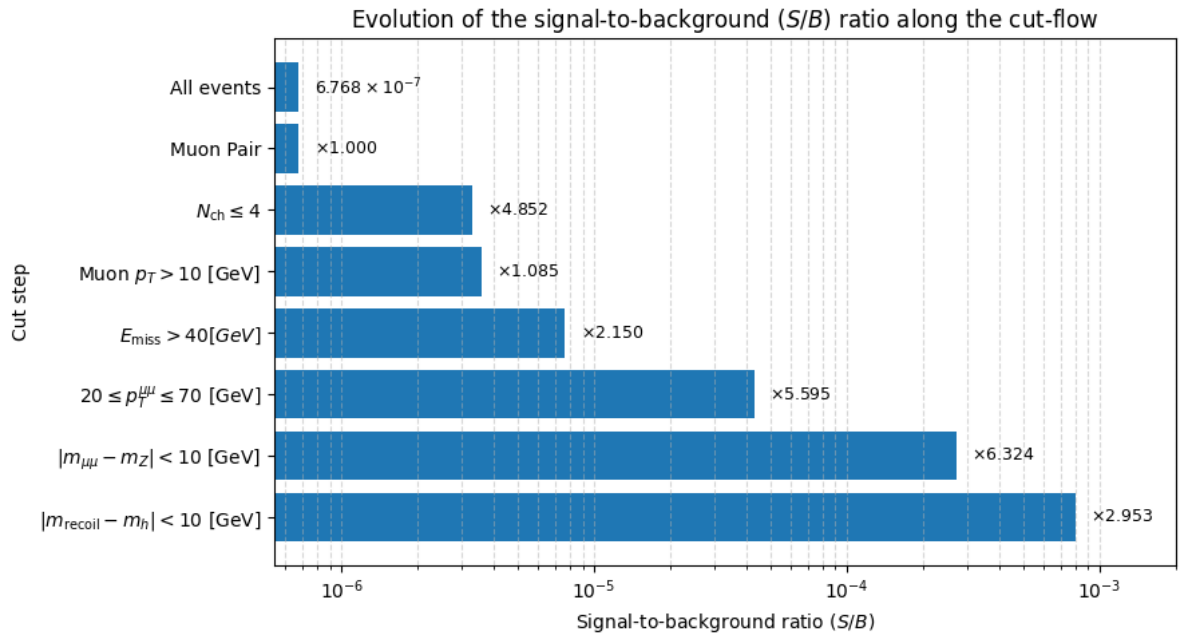


Figure 33 – Evolution of the signal-to-background ratio (S/B) along the cutflow. The y-axis has the cuts done, starting from the bottom to the top, the x-axis is the absolute value of the S/B ratio in log. The starting absolute S/B value before each selection is shown in the first bar. Subsequent annotations indicate the ratio of S/B relative to the previous cut, highlighting the impact of each selection on the sample purity.

the squared relative uncertainty for $\mathrm{BR}(h \to \mathrm{inv})$ can then be written for:

$$\left(\frac{\delta \mathrm{BR}(h \to \mathrm{inv})}{\mathrm{BR}(h \to \mathrm{inv})}\right)^2 = \left(\frac{\delta \sigma_{\mathrm{signal}}}{\sigma_{\mathrm{signal}}}\right)^2 + \left(\frac{\delta \sigma(Zh)}{\sigma(Zh)}\right)^2 + \left(\frac{\delta \mathrm{BR}(Z \to \mu^+\mu^-)}{\mathrm{BR}(Z \to \mu^+\mu^-)}\right)^2 \tag{5.8}$$

the relative errors of this formula can be split into two categories, σ_{signal} is defined as statistical because its values depends on the results of the MC samples and $\sigma(Zh)$ and $\mathrm{BR}(Z \to \mu^+\mu^-)$ as systematic, since their values are defined *a priori*. All yields are normalized to an integrated luminosity of $\mathcal{L} = 250\ \mathrm{fb}^{-1}$, assuming the ILC capability of polarizing beams for $e_L^- e_R^+$ and the signal cross-section described in Section 3.5.2. The signal is the IHD, defined in Section 2.3.1, at the ILC, the *Higgstrahlung* Zh production cross-section at $\sqrt{s} = 250\,\mathrm{GeV}$ with polarized beams, $P(e^-) = -0.8$ and $P(e^+) = +0.3$.

The combination of all these values leads to the calculation of the effective signal cross-section is therefore, rewriting equation 3.8, allows to calculate the initial expected signal yield:

$$S_{\mathrm{expected}} = N_{\mathrm{gen}} \times w_{\mathrm{signal}} = \sigma_{\mathrm{signal}} \times \mathcal{L} \tag{5.9}$$

$$= 1.05 \times 10^{-2} \times 250 \tag{5.10}$$

$$= 2.63 \tag{5.11}$$

thus, approximately three signal events are expected prior to any event selection. The initial signal Monte Carlo sample contains $N_{\text{gen}} = 2388$ generated events. Each event is assigned a weight defined in 3.8 which yielded $w_{\text{signal}} = 1.103 \times 10^{-3}$. This normalization ensures that the weighted event yields correspond to the expected number of events for the assumed luminosity. After all cuts, the number of remaining signal samples was:

$$N_{sel} = 2151 \tag{5.12}$$

therefore the MC statistical error for signal derived in equation 3.19 is equal to:

$$\sigma_{\text{stat}} = \sqrt{N_{\text{sig}}} \cdot w = \sqrt{2151} \times 1.103 \times 10^{-3} \tag{5.13}$$

$$= 5.11 \times 10^{-2} \tag{5.14}$$

now computing the relative MC statistical uncertainty:

$$\frac{\sigma_{\text{stat}}}{\sigma(Zh)} = \frac{5.11 \times 10^{-2}}{2.63} = 2.01\% \tag{5.15}$$

the relative uncertainty for the *Higgstrahlung* $\sigma(Zh)$ cross-section provided by (4) is:

$$\frac{\delta\sigma(Zh)}{\sigma(Zh)} = 2.5\% \tag{5.16}$$

now the relative uncertainty for $\text{BR}(Z \to \mu^+\mu^-)$ provided by (14) is:

$$\frac{\delta\text{BR}(Z \to \mu^+\mu^-)}{\text{BR}(Z \to \mu^+\mu^-)} = \frac{6.6 \times 10^{-4}}{3.366 \times 10^{-2}} = 1.96\% \tag{5.17}$$

the combination of these two values results in thesystematic uncertainty for $\sigma(Zh)$:

$$\frac{\sigma_{\text{sys}}}{\sigma(Zh)} = \sqrt{(0.025)^2 + (0.0196)^2} = 3.18\% \tag{5.18}$$

now combining the systematic and statistical errors, the total relative uncertainty on $\text{BR}(h \to \text{inv})$ is:

$$\frac{\delta\text{BR}(h \to \text{inv})}{\text{BR}(h \to \text{inv})} \approx 3.8\% \tag{5.19}$$

Within the limited set of propagated uncertainties considered here, the result appears to be dominated by external normalization inputs. It is important to emphasize that this estimate only propagates external normalization uncertainties on $\sigma(Zh)$ and $\text{BR}(Z \to \mu^+\mu^-)$, together with Monte Carlo statistical fluctuations. No dedicated detector-level or background modeling systematic uncertainties presented at Table 20 are evaluated in this study. Therefore, the quoted uncertainty should be interpreted as an illustrative projection rather than a complete systematic assessment.

5.3 Implementation of the Analysis Framework

This short section is summary of the workflow done, starting with the Key4hep implementation to data analysis, divided into three subsections.

Table 20 – Relative uncertainty contributions to the extraction of BR($H \to$ inv) at $\mathcal{L} = 250\,\text{fb}^{-1}$. Only Monte Carlo statistical and external normalization (parametric) uncertainties are included. All contributions are combined in quadrature.

Source	Relative Uncertainty (%)	Contribution Type
Signal MC statistics	2.01	Statistical
$\sigma(Zh)$ projection	2.50	Systematic
BR($Z \to \mu^+\mu^-$)	1.96	Systematic
Combined systematic	3.18	–
Total uncertainty	3.76	–

5.3.1 Data Layer – Key4hep Steering File

Key4hep is easily integrated and scalable, the construction of Table 28 was done with the combination of the Producion Summary and ELOG gendata for all `eL.pR` with a Center-of-Mass energy of $\sqrt{s} = 250\,\text{GeV}$, allowing to have easy access to all `process_id` and their respective input `slcio` files in `250-SetA`, the signal input `ROOT` files were provided separately. Since all processes were simulated separately, this meant that a `ROOT` file was produced for each process and allowing easy distinction between samples origin. The biggest pitfalls in the steering file was the necessity of using two `MarlinProcessors`, `Output_DST` and `LCIOWriter`, both which only had the purpose of serving as and intermediary to translate `slcio` into a readable form for the `MCConsumer` Gaudi algorithm, after that they produced two `slcio` files with an average size of $\approx$ 140MB with no utility for the data analysis.

5.3.2 Data Layer – Gaudi as a Physics Reconstruction Interface

The only criteria for the `MCConsumer` algorithm to decide which event was saved in the `ROOT` file was the presence of $\mu^+\mu^-$:

```
...
if (mu1.charge * mu2.charge >= 0) {
// same-sign or undefined charge
return StatusCode::SUCCESS;
}
...
```

this means that every event that didn't fit this criteria only contributed to the N_{gen} value. For events that did fit into the criteria, the invariant and recoil mass were calculate inside the algorithm:

```
...
// Total 4-momentum of the muon system
TLorentzVector dimuon = p4_1 + p4_2;
m_invMass = (p4_1 + p4_2).M();
...
```

```
    // Initial 4-momentum of e+e- system
    TLorentzVector initial(0, 0, 0, m_ecm);
    // Recoil 4-momentum
    TLorentzVector recoil = initial - dimuon;
    m_recoilMass = recoil.M();
    ...
```

the `M()` function has functionality of the calculating the squared root of the module of any vector, resulting in the equations 4.22 and 4.33.

5.3.3 Statistical Evaluation Layer

An implementation of a `bash` script that accepts custom arguments to run simulations for each process, producing a distinct `ROOT` output file per process. This workflow allowed for easy reproducibility and modularity as new processes could be integrated or removed from the workflow without heavy manual work. With the variables produced in Table 13, the data analysis: weights calculation, weighted yields, graphs and sensibility estimation was done outside the Key4hep ecosystem using mostly `python` libraries like `numpy`, `pandas`, `uproot` and etc.

5.4 Final Signal and Background Composition

The cutflow Tables 17 and 18 shows the improvement in the signal isolation sequentially with Table 21. The IHD is rare event, therefore, it's expected that the FoM and signal-to-background yield S/B values to be low. Since the signal-to-background ratio evolves multiplicatively along the cutflow and spans more than one order of magnitude, for the Figure 33, a logarithmic scale was employed in the x-axis to better visualize relative improvements. The initial signal yield is calculated based on the number of generated events $N = 2388$, this leads $S_i = 2.63$, after all cuts, the final signal yield was $S_f = 1.66$. Therefore, the final efficiency ϵ was:

$$\epsilon = \frac{1.66}{2.63} = 0.631 \tag{5.20}$$

this means the signal efficiency is approximately $\epsilon \approx 63.1\%$, more than half of the initial samples remained.

The following sections present two complementary sensitivity projections. First, Section 5.4.1, an approximate statistical sensitivity is estimated using a Gaussian counting approach. Second, Section 5.4.2, the impact of background normalization uncertainties is evaluated to assess the stability of the projected sensitivity within this simplified framework.

Table 21 – Signal yield S, background yield B, total yield $S + B$, signal-to-background S/B and statistical sensitivity after the full selection. The results correspond to $\mathcal{L} = 250\,\text{fb}^{-1}$ and $\text{BR}(h \to \text{inv}) = 10^{-3}$.

Quantity	Value
Signal yield S	1.665
Background yield B	2079.498
Total yield $S + B$	2081.163
Signal-to-background $\frac{S}{B}$	8.0×10^{-4}
Statistical sensitivity $\frac{S}{\sqrt{S+B}}$	3.6508×10^{-2}

5.4.1 Approximate Counting-Based Sensitivity Estimate

The signal-to-background ratio $S/B \approx 8.0 \times 10^{-4}$, available at Table 21, indicates that the samples are at a strongly background-dominated regime $S \ll B$. Using the final signal (S) and background (B) yields from Table 21, a simple counting-based estimate of the statistical sensitivity is performed. In this regime, the significance can be approximated by

$$Z_0 = \frac{S}{\sqrt{B}} = \frac{1.665}{\sqrt{2079.5}} = 0.0365. \tag{5.21}$$

for the reference BR $\text{BR}_0 = 10^{-3}$. Since the signal yield scales linearly with $\text{BR}(h \to \text{inv})$, the approximate significance behaves as

$$Z(\text{BR}) = Z_0 \cdot \frac{\text{BR}}{\text{BR}_0}. \tag{5.22}$$

Under the Gaussian approximation valid for large event counts, a one-sided 95% confidence criterion corresponds to $Z = 1.64$. Using this value as a sensitivity benchmark, the BR required to reach this significance is

$$\text{BR}(h \to \text{inv})_{Z=1.64} = 0.001 \times \frac{1.64}{0.0365} \approx 4.5\%. \tag{5.23}$$

This value represents the BR that would yield a counting-based significance of $Z = 1.64$ under Gaussian assumptions. No likelihood construction, nuisance parameter treatment, or CL_s evaluation is performed. Therefore, this number should be interpreted purely as an indicative statistical sensitivity scale at $\mathcal{L} = 250\,\text{fb}^{-1}$ rather than a formal 95% confidence level exclusion.

For the benchmark value $\text{BR}(h \to \text{inv}) = 10^{-3}$, the resulting statistical significance remains small, as expected for such a suppressed branching fraction. Within this simplified counting framework, $\text{BR}(h \to \text{inv})$:

$$S \propto \mathcal{L}, \quad B \propto \mathcal{L} \tag{5.24}$$

implying that the significance is proportional to:

$$Z \propto \sqrt{\mathcal{L}} \tag{5.25}$$

and consequently

$$\mathrm{BR}(h \to \mathrm{inv})_{Z=1.64} \propto \frac{1}{\sqrt{\mathcal{L}}} \tag{5.26}$$

This scaling illustrates the expected statistical improvement with increasing integrated luminosity within the counting approximation. A full statistical interpretation would require construction of a profile likelihood including systematic nuisance parameters, which is beyond the scope of this work.

5.4.2 Impact of Relative Background Normalization Uncertainty

To assess the robustness of the counting-based sensitivity estimate, a simplified study of background normalization uncertainty is performed. Rather than constructing a full likelihood with nuisance parameters, the effect is modeled phenomenologically by inflating the variance of the background yield.

Analogous to the previous section, the regime $S \ll B$ can be assumed, the counting-based significance can be approximated by

$$Z(\delta_B) \approx \frac{S}{\sqrt{B + (\delta_B B)^2}} \tag{5.27}$$

where δ_B denotes the assumed fractional uncertainty on the total background normalization. This expression corresponds to adding a normalization uncertainty in quadrature to the Poisson variance of the background $\sigma_B^2 = B + (\delta_B B)^2$, thereby providing a simplified estimate of its impact on sensitivity.

A degradation factor $D(\delta_B)$ is introduced to evaluate the evolution between the significances for each uncertainty, for each relative uncertainty estimation:

$$D(\delta_B) = \frac{Z(0)}{Z(\delta_B)} = \frac{\sqrt{B + (\delta_B B)^2}}{\sqrt{B}}, \tag{5.28}$$

the initial $Z_{0\%}$ is divided by the resulting $Z_{X\%}$ to get the degradation factor, which describes how the counting-based sensitivity changes with increasing background uncertainty.

Using the same Gaussian benchmark ($Z = 1.64$) introduced in the previous section, and since the significance is directly proportional to the BR $S \propto \mathrm{BR}(h \to \mathrm{inv})$, rewriting equation 5.28, results in the relation:

$$\mathrm{BR}(h \to \mathrm{inv})_{Z=1.64}(\delta_B) = \mathrm{BR}(h \to \mathrm{inv})_{Z=1.64}(0) \times D(\delta_B). \tag{5.29}$$

This relation should be interpreted as a scaling estimate within the counting approximation, not as the result of a profile likelihood or CL_s procedure. In the regime where the systematic term dominates, $\delta_B B \gg \sqrt{B}$, the expression 5.27 simplifies to

$$Z \approx \frac{S}{\delta_B B}. \tag{5.30}$$

showing the direct relation between the BR CL and the degradation factor. Solving for the signal yield required to reach the Gaussian benchmark $Z = 1.64$ gives

$$S_{Z=1.64} \approx 1.64 \times \delta_B B \tag{5.31}$$

In this limit, the achievable sensitivity becomes directly proportional to the fractional background uncertainty. Now, sensitivity scales linearly with B, meaning:

$$\mathrm{BR}(h \to \mathrm{inv})_{\mathrm{sensitivity}} \propto \delta_B \tag{5.32}$$

and no longer improves with increasing luminosity. This behavior illustrates the transition from a statistically limited regime to a normalization-systematics-limited regime within the simplified counting framework.

5.4.2.1 No Background Normalization Uncertainty ($\delta_B = 0\%$)

In the absence of background normalization uncertainty, the effective variance reduces to the Poisson expectation,

$$\sqrt{B + (\delta_B B)^2} = \sqrt{B} = \sqrt{2079.4} = 45.60. \tag{5.33}$$

Using the Gaussian benchmark $Z = 1.64$, the signal yield required to reach this sensitivity is

$$S^{0\%}_{Z=1.64} = 1.64 \times 45.60 = 74.79. \tag{5.34}$$

In this purely statistical regime, the sensitivity is entirely driven by Poisson fluctuations and represents the optimal performance within the counting approximation.

5.4.2.2 5% Background Normalization Uncertainty ($\delta_B = 5\%$)

Assuming a fractional background uncertainty of 5%,

$$\sqrt{B + (\delta_B B)^2} = \sqrt{2079.4 + (0.05 \times 2079.4)^2} = 113.53. \tag{5.35}$$

The corresponding signal yield required to reach the Gaussian benchmark becomes

$$S^{5\%}_{Z=1.64} = 1.64 \times 113.53 = 186.20. \tag{5.36}$$

The systematic contribution $\delta_B B \approx 104$ already exceeds the statistical fluctuation $\sqrt{B} \approx 46$, indicating the onset of a normalization-dominated regime.

The relative degradation with respect to the statistical-only case is

$$D(5\%) = \frac{113.53}{45.60} = 2.49. \tag{5.37}$$

Thus, a 5% background normalization uncertainty increases the required signal yield by a factor of 2.49 within the counting-based sensitivity estimate.

5.4.2.3 10% Background Normalization Uncertainty ($\delta_B = 10\%$)

For a 10% fractional background uncertainty,

$$\sqrt{B + (\delta_B B)^2} = \sqrt{2079.4 + (0.10 \times 2079.4)^2} = 212.89. \tag{5.38}$$

The signal yield required to reach the Gaussian benchmark is

$$S^{10\%}_{Z=1.64} = 1.64 \times 212.89 = 349.14. \tag{5.39}$$

In this regime, $\delta_B B \gg \sqrt{B}$, systematic contribution $\delta_B B \approx 208$ is four times bigger than $\sqrt{B}$, and the sensitivity becomes dominated by the normalization uncertainty.

The corresponding degradation factor is

$$D(10\%) = \frac{212.89}{45.60} = 4.67. \tag{5.40}$$

This demonstrates the transition from a statistically limited regime to a normalization-systematics-limited regime within the simplified counting framework.

5.4.3 Regime Summary

Sections 5.4.1 and 5.4.2 illustrate, within the simplified counting-based framework, the transition between two sensitivity regimes:

- Statistics-dominated regime: $\delta_B B \ll \sqrt{B}$, where the sensitivity improves proportionally to $\sqrt{\mathcal{L}}$,

- Normalization-systematics-dominated regime: $\delta_B B \gg \sqrt{B}$, where the achievable sensitivity becomes limited by background normalization uncertainty and no longer significantly benefits from increased luminosity.

The present study shows that normalization uncertainties at the few-percent level already push the analysis toward the systematics-dominated regime within this counting approximation, values displayed at Table 22.

Table 22 – Projected degradation of the counting-based branching ratio sensitivity benchmark ($Z = 1.64$) as a function of the assumed fractional background normalization uncertainty δ_B. The required signal yield $S_{Z=1.64}$ is computed using the Gaussian counting approximation with statistical and normalization contributions added in quadrature. The degradation factor $D(\delta_B)$ quantifies the relative worsening of the sensitivity with respect to the purely statistical scenario. No profile likelihood or CL_s procedure is performed.

δ_B	$\mathrm{BR}(h \to \mathrm{inv})_{Z=1.64}$	$S_{Z=1.64}$	$D(\delta_B)$
0%	4.5%	74.79	1.00
5%	11.2%	186.20	2.49
10%	21.0%	349.14	4.67

5.5 Contextualization of the Projected Sensitivity

The projected sensitivity obtained in this analysis is broadly consistent with previous studies of IHDs at the ILC and the LHC (1, 2, 4, 3, 5, 32). In particular, ILC Higgs recoil analyses at $\sqrt{s} = 250\,\mathrm{GeV}$ with integrated luminosities up to $\mathcal{L} = 900\ \mathrm{fb}^{-1}$, and at $\sqrt{s} = 500\,\mathrm{GeV}$ with $\mathcal{L} = 1600\ \mathrm{fb}^{-1}$, typically report sensitivities at the percent level.

For comparison, the current upper limit reported by the Particle Data Group is $\mathrm{BR}(h \to \mathrm{inv}) < 10.7\%$ (14). Several ILC benchmark studies assume $\mathrm{BR}(h \to \mathrm{inv}) = 10\%$ when estimating performance (2, 5). In contrast, this thesis assumes $\mathrm{BR}(h \to \mathrm{inv}) = 0.1\%$, which is two orders of magnitude smaller. Since the signal cross-section scales linearly with the invisible BR, $\sigma_{\mathrm{signal}} \propto \mathrm{BR}(h \to \mathrm{inv})$, the expected signal yield scales accordingly.

Differences with respect to previous projections are therefore expected. First, the signal yield depends directly on both the assumed BR, $S \propto \mathrm{BR}(h \to \mathrm{inv})$, and the integrated luminosity, $S \propto \mathcal{L}$. Second, several benchmark ILC studies employ multivariate techniques such as Boosted Decision Trees (BDTs) to enhance signal–background separation (2). These methods can reduce the background yield by an additional order of magnitude relative to a purely cut-based analysis. Further variations arise from different detector modeling assumptions, the treatment of ISR and *beamstrahlung* effects, and the inclusion of systematic uncertainties.

It is therefore important to emphasize that the present result should be interpreted as a simplified cut-based sensitivity estimate within the Key4hep framework. No full profile-likelihood fit, multivariate optimization, or detailed systematic uncertainty model has been implemented. Consequently, the projected sensitivity should be viewed as a proof-of-principle conservative baseline rather than a fully optimized performance projection.

Overall, the obtained sensitivity lies within the expected envelope of ILC Higgs recoil projections at $\sqrt{s} = 250\,\mathrm{GeV}$, demonstrating the robustness of the recoil technique even within a simplified framework.

6 Conclusions and Outlook

This work presents a workflow and proof-of-principle study of the IHD through the Zh recoil method with $Z \to \mu^+\mu^-$ at the at $\sqrt{s} = 250\,\text{GeV}$ within the Key4hep ecosystem. Rather than aiming at a fully optimized or ultimate sensitivity result, the focus has been on establishing the Key4hep ecosystem as a robust, reproducible, modular, flexible and scalable framework, combining with analysis chain that demonstrates the feasibility of the approach and provides a solid foundation for future refinements and optimization studies.

Key4hep is a modern analysis framework designed to provide machine-agnostic measurements, an especially relevant feature given that the International Linear Collider will operate with two detectors, the International Large Detector (ILD) and the Silicon Detector (SiD). The framework demonstrates the capability of handling workflows based on both `ROOT` and `slcio` data formats, enabling flexible analysis configurations.

Nevertheless, Key4hep still presents practical limitations. Its documentation does not yet fully expose the breadth of its capabilities, which may hinder new users. In addition, while `ROOT` files can be processed natively, handling `slcio` inputs still requires the use of legacy components such as `MarlinProcessors` to bridge the data into the `EDM4hep` format.

Improved documentation and further integration of legacy data formats would significantly facilitate the migration from previous frameworks and strengthen Key4hep's position as the standard analysis ecosystem for future collider studies.

A simple modification of the code could include the other beams polarizations available at Table 5 or values for $\sqrt{s}$ though the introduction of custom arguments for the `k4run` commands, such as `--eX.pY` where `X` and `Y` could be any polarization and `--sZZZ` where `ZZZ` could br the Center-of-Mass energy. More Gaudi algorithms can also be introduced by simply adding into the `components` folders and importing them in the `python` steering file.

Beyond software considerations, statistical limitations also impact the present study. A limiting factor in the small sample size for the signal was due the fact that generated data is limited to the generator files, the input files. This means that only the same simulated samples could be produced, however, this is what allows reproducibility of this study. Being able to increase the total data produced, including the background, would significantly increase the diversity for the signal yield.

Future developments can be organized into three main directions: improvements in the statistical framework, incorporation of realistic systematic effects, and optimization

of the event selection strategy.

From a statistical perspective, the current cut-based significance estimate should be replaced by a profile-likelihood framework, even in a simplified single-bin implementation, including explicit nuisance parameters for dominant systematic uncertainties. This would ensure a consistent treatment of both normalization and shape effects, elevating the present proof-of-principle analysis into a fully fledged sensitivity study.

Beyond statistical formalism, several sources of systematic uncertainty must be incorporated. Beam-related effects such as ISR and *beamstrahlung* can distort the recoil mass distribution and modify the visible multiplicity, as suggested by the N_{ch} behavior shown in Figure 16. Alternative generator configurations could be used to quantify these distortions and propagate their impact to the signal extraction. Similarly, detector-driven uncertainties from the International Large Detector (ILD) and Silicon Detector (SiD), particularly muon momentum scale and resolution variations, should be propagated through the reconstruction chain to evaluate their effect on the recoil peak stability.

Finally, the event selection strategy itself can be optimized. The recoil mass window and associated kinematic requirements could be re-optimized using an objective metric such as the expected confidence level limit. Alternatively, the cut-based approach may be replaced by multivariate analysis techniques trained on dominant background processes. Machine Learning models, including Boosted Decision Trees as explored in previous studies (2), can exploit correlations between multiple observables beyond the cutflow paradigm. While such models are often criticized for their "black box" behavior, recent developments in physics-inspired neural networks aim to preserve interpretability by embedding physically motivated constraints into the training procedure. The combination of multiple Z decay channels, such as $Z \to e^+e^-$ and potentially hadronic modes, within a unified likelihood framework would further extend the achievable branching ratio sensitivity.

In conclusion, the workflow established in this thesis provides a reproducible foundation upon which a full-scale sensitivity program can be constructed once larger simulated datasets and refined systematic treatments become available.

Bibliography

1 FUJII, K. et al. *Physics Case for the 250 GeV Stage of the International Linear Collider*. arXiv, 2018. ArXiv:1710.07621 [hep-ex]. Disponível em: <http://arxiv.org/abs/1710.07621>. Cited 10 times in pages 11, 13, 33, 40, 43, 44, 46, 47, 50, and 107.

2 POTTER, C. et al. *Expected Sensitivity to Invisible Higgs Boson Decays at the ILC with the SiD Detector (A Snowmass White Paper)*. arXiv, 2022. ArXiv:2203.08330 [hep-ph]. Disponível em: <http://arxiv.org/abs/2203.08330>. Cited 22 times in pages 11, 13, 21, 22, 33, 46, 47, 55, 61, 62, 63, 69, 71, 73, 74, 75, 86, 88, 95, 107, 110, and 125.

3 GHOSH, D. et al. Looking for an invisible Higgs signal at the LHC. *Physics Letters B*, v. 725, n. 4-5, p. 344–351, out. 2013. ISSN 03702693. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0370269313006060>. Cited 12 times in pages 11, 13, 40, 41, 42, 43, 55, 73, 74, 86, 95, and 107.

4 YAN, J. et al. Measurement of the {Higgs} boson mass and {$e^+e^- \to zh$} cross section using {$z \to \mu^+\mu^-$} and {$z \to e^+ e^-$} at the {ILC}. *Physical Review D*, v. 94, n. 11, p. 113002, dez. 2016. ISSN 2470-0010, 2470-0029. ArXiv:1604.07524 [hep-ex]. Disponível em: <http://arxiv.org/abs/1604.07524>. Cited 11 times in pages 11, 13, 21, 33, 59, 63, 73, 88, 95, 100, and 107.

5 KATO, Y. *Probing the dark sector via searches for invisible decays of the Higgs boson at the ILC*. arXiv, 2020. ArXiv:2002.12048 [hep-ex]. Disponível em: <http://arxiv.org/abs/2002.12048>. Cited 10 times in pages 11, 13, 33, 47, 61, 73, 74, 88, 95, and 107.

6 PANDUROVIć, M. *Physics potential for the measurement of {\large ${\sigma(HZ)\times BR(H \rightarrow WW^{*})}$} at the 250 GeV ILC*. arXiv, 2019. ArXiv:1902.08032 [hep-ex]. Disponível em: <http://arxiv.org/abs/1902.08032>. Cited 4 times in pages 11, 13, 73, and 95.

7 Wikipedia contributors. *Standard Model — Wikipedia, The Free Encyclopedia*. 2026. Disponível em: <https://en.wikipedia.org/w/index.php?title=Standard_Model&oldid=1351551687>. Cited 2 times in pages 15 and 39.

8 BAER, H. et al. *The International Linear Collider Technical Design Report - Volume 2: Physics*. arXiv, 2013. ArXiv:1306.6352 [hep-ph]. Disponível em: <http://arxiv.org/abs/1306.6352>. Cited 4 times in pages 15, 40, 44, and 51.

9 DJOUADI, A. The Anatomy of Electro-Weak Symmetry Breaking. I: The Higgs boson in the Standard Model. *Physics Reports*, v. 457, n. 1-4, p. 1–216, fev. 2008. ISSN 03701573. ArXiv:hep-ph/0503172. Disponível em: <http://arxiv.org/abs/hep-ph/0503172>. Cited 4 times in pages 15, 37, 45, and 47.

10 International Linear Collider. *International Linear Collider*. 2026. Disponível em: <https://linearcollider.org/>. Cited 2 times in pages 15 and 49.

11 BEHNKE, T. et al. *The International Linear Collider Technical Design Report - Volume 4: Detectors*. arXiv, 2013. ArXiv:1306.6329 [physics]. Disponível em: <http://arxiv.org/abs/1306.6329>. Cited 8 times in pages 15, 21, 50, 51, 52, 54, 55, and 63.

12 PICH, A. Electroweak Symmetry Breaking and the Higgs Boson. *Acta Physica Polonica B*, v. 47, n. 1, p. 151, 2016. ISSN 0587-4254, 1509-5770. ArXiv:1512.08749 [hep-ph]. Disponível em: <http://arxiv.org/abs/1512.08749>. Cited 3 times in pages 21, 45, and 47.

13 KAWADA, S.-i. *The mini-DST: a high-level LCIO format*. arXiv, 2021. ArXiv:2105.08622 [physics] version: 1. Disponível em: <http://arxiv.org/abs/2105.08622>. Cited 2 times in pages 22 and 118.

14 NAVAS, S.; others. Review of particle physics. *Phys. Rev. D*, v. 110, n. 3, p. 030001, 2024. Cited 5 times in pages 22, 61, 100, 107, and 125.

15 ILD ELOG: Generator Meta Data (genmeta). Niigata University / ILD Collaboration, 2025. Published: ELOG Web Logbook for ILD Monte Carlo Production, International Large Detector (ILD) Collaboration. Disponível em: <https://ild.ngt.ndu.ac.jp/elog/genmeta/>. Cited 4 times in pages 23, 60, 61, and 129.

16 AAD, G. et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, Elsevier, v. 716, n. 1, p. 1–29, 2012. Cited in page 33.

17 BASS, S. D.; ROECK, A. D.; KADO, M. The Higgs boson – its implications and prospects for future discoveries. *Nature Reviews Physics*, v. 3, n. 9, p. 608–624, jul. 2021. ISSN 2522-5820. ArXiv:2104.06821 [hep-ph]. Disponível em: <http://arxiv.org/abs/2104.06821>. Cited 5 times in pages 33, 37, 38, 45, and 46.

18 BAMBADE, P. et al. *The International Linear Collider: A Global Project*. arXiv, 2019. ArXiv:1903.01629 [hep-ex]. Disponível em: <http://arxiv.org/abs/1903.01629>. Cited in page 33.

19 GANIS, G.; HELSENS, C.; VöLKL, V. *Key4hep, a framework for future HEP experiments and its use in FCC*. arXiv, 2021. ArXiv:2111.09874 [hep-ex]. Disponível em: <http://arxiv.org/abs/2111.09874>. Cited 2 times in pages 34 and 117.

20 ABOKHALIL, A. The Higgs Mechanism and Higgs Boson: Unveiling the Symmetry of the Universe. *arXiv preprint arXiv:2306.01019*, 2023. Cited 4 times in pages 37, 39, 43, and 46.

21 ENGLERT, F.; HIGGS, P. W. *The Nobel Prize in Physics 2013: Summary*. 2013. Published: NobelPrize.org. Disponível em: <https://www.nobelprize.org/prizes/physics/2013/summary/>. Cited in page 37.

22 LARKOSKI, A. J. *Elementary particle physics: an intuitive introduction*. [S.l.]: Cambridge University Press, 2019. Cited in page 43.

23 votatera. *Yukawa Coupling | HEP Concepts, Quark Mass & Interaction*. 2024. Published: Modern-Physics.org. Disponível em: <https://modern-physics.org/yukawa-coupling/>. Cited in page 44.

24 COLLABORATION, C. M. S. Search for Higgs boson decays into Z and J/ψ and for Higgs and Z boson decays into J/ψ or Υ pairs in pp collisions at $\sqrt{s}=13$ TeV. *Physics Letters B*, v. 842, p. 137534, jul. 2023. ISSN 03702693. ArXiv:2206.03525 [hep-ex]. Disponível em: <http://arxiv.org/abs/2206.03525>. Cited in page 46.

25 ALLDAY, J. *Quarks, leptons and the big bang*. [S.l.]: CRC Press, 2016. Cited in page 53.

26 ALVES, G. A.; HENSEL, C. *Estudos de triggers de energia transversa perdida e vértices secundários para uma busca por Supersimetria em estados finais com jatos, energia transversa perdida e vértices secundários & Contribuições para o calorímetro hadrônico frontal do experimento CMS/LHC*. Tese (Doutorado) — Brazilian Center for Research in Physics, Brazilian Center for Research in Physics, 2018. Cited in page 53.

27 COLLABORATION, L.; CLOSIER, J. *VErtex LOcator (VELO)*. 2026. Published: LHCb Outreach, CERN. Disponível em: <https://lhcb-outreach.web.cern.ch/detector/vertex-locator-velo/>. Cited in page 55.

28 STANDALONE conversion from LCIO to EDM4hep. Published: Key4hep Documentation, Key4hep Project, CERN. Disponível em: <https://key4hep.github.io/key4hep-doc/main/how-tos/k4edm4hep2lcioconv/doc/LCIO2EDM4hep.html>. Cited in page 56.

29 ILD MC-2020 Production Summary. Niigata University / ILD Collaboration, 2021. Published: MC-2020 Monte Carlo Production Monitoring Page, International Large Detector (ILD). Disponível em: <https://ild.ngt.ndu.ac.jp/mc-prod/prodmon/prodsum-mc2020.html>. Cited in page 60.

30 MADLENER, T. *Replace an Existing Marlin Processor with a Gaudi (Functional) Algorithm*. [S.l.]: DESY, 2025. Cited in page 65.

31 FEICHTINGER, P. et al. Punzi-loss: A non-differentiable metric approximation for sensitivity optimisation in the search for new particles. *The European Physical Journal C*, v. 82, n. 2, p. 121, fev. 2022. ISSN 1434-6044, 1434-6052. ArXiv:2110.00810 [hep-ex]. Disponível em: <http://arxiv.org/abs/2110.00810>. Cited in page 71.

32 EBOLI, O. J. P.; ZEPPENFELD, D. Observing an invisible Higgs boson. *Physics Letters B*, v. 495, n. 1-2, p. 147–154, dez. 2000. ISSN 03702693. ArXiv:hep-ph/0009158. Disponível em: <http://arxiv.org/abs/hep-ph/0009158>. Cited in page 107.

33 GAEDE, F. et al. EDM4hep - a common event data model for HEP experiments. *PoS*, ICHEP2022, p. 1237, nov. 2022. Cited in page 120.

Appendix

APPENDIX A – Key4hep Framework

The analysis presented in this work was performed within the *Key4hep* software stack (19), a turnkey and modular framework designed for simulation, reconstruction, and analysis studies at future particle colliders such as the ILC and FCC[1].

Key4hep provides a coherent environment that integrates detector simulation, event reconstruction, and physics analysis tools within a common infrastructure. Its main components can be grouped according to their functionality:

A.0.1 Core Infrastructure

- `Gaudi` – Data processing framework that manages the execution flow of algorithms.
- Build Infrastructure – CMake-based system ensuring consistent compilation and deployment.

Detector Description and Simulation

- `DD4hep` – Toolkit for detector geometry description and conditions handling.
- `Geant4` – Particle transport and detector response simulation.

Event Data Model and I/O

- `PODIO` – Event data model toolkit used to define and persist data structures.

Analysis Tools

- `ROOT` – Statistical analysis, histogramming, and visualization framework.

A.1 Simulation Pipeline

All the code used is located at the author's github[2]. Key4hep allows the usage of `python` script, named `run_example.py`, to function as the main driver of the simulation. DESY computers have a variety of available version, using the following command, the desired version is set:

[1] Documentation available at <https://key4hep.github.io/key4hep-doc/>.
[2] <https://github.com/mazeeqe/tomaz-example>

```
source /cvmfs/sw.hsf.org/key4hep/setup.sh -r 2025-05-29
```

Listing A.1 – Source key4hep command.

this loads the version of the date `2025-05-29`.

```
source setup.sh
```

After loading the required version, change to the build directory:

```
cd build
```

Listing A.2 – Changing to build folder.

Then, through the terminal, running the following command will do the simulation:

```
k4run ../example/options/run_example.py
```

Listing A.3 – key4hep command.

An additional argument is available to split between signal, `-signal`, and background simulation, `-XXXXXX`, where `XXXXX` is an process id, the full list of available ids is present at annex C, this allows an easier management of the output data. To run each desired option:

```
k4run ../example/options/run_example.py --signal
k4run ../example/options/run_example.py --XXXXXX
```

Listing A.4 – Commands for Signal and Background simulation respectively.

with no arguments, the signal option is default. Inside the python script the collections are selected.

```
collections = ["PandoraPFOs", "EventHeader", "MCParticlesSkimmed"]
```

Listing A.5 – Collections used in the code.

each collection has a type (13), `PandoraPFOs` is used in the `MCConsumerAlg` Gaudi algo-

Table 23 – Gaudi Collections used and their purpose. Adaptation of table (1) for the contents of mini-DST from (13).

Collection Name	Collection Type	Explanation
`PandoraPFOs`	ReconstructedParticle	particle flow objects
`EventHeader`		

rithm,

```
consumer = MCConsumerAlg("MCConsumer")
consumer.RecoParticleColl = "PandoraPFOs"
```

An `bash` script is available that does the simulations for the signal and all process ids. Within the `build` folder, run:

```
chmod +x ../scripts/init.sh
```

this will turn the `init.sh` into a functional bash script, then do:

```
../scripts/init.sh
```

after each simulated process, the `ROOT` files produced are moved to their respective folder in `output_files`.

A.2 Background

Key4hep does not have native support for `slcio` and since background inputs are `slcio` files. The conversion is done with

```
MarlinProcessorWrapper("Output_DST")
```

the "Output_DST `Lcio` and `EDM4hep` have different names for their collections. A mapping is needed to convert, the `Lcio2EDM4hepTool` tool does the conversion:

```
        lcio2edm = Lcio2EDM4hepTool("Lcio2EDM4hepTool")
        lcio2edm.collNameMapping = {
        "MCParticles": "MCParticles",
        "PandoraPFOs": "PandoraPFOs"
        }
```

A.3 Application Manager

The `ApplicationMgr` is what bridges the input files, the Gaudi algorithm and the external services:

```
ApplicationMgr(
    # provide list and order of algorithms
    TopAlg=algs,
    EvtSel="NONE",
    EvtMax=events_simulated,
    ExtSvc=[io_svc],
    OutputLevel=INFO
)
```

`TopAlg` is the list of Gaudi algorithms that will be executed during the simulation, in this case, only the `MCConsumer` is in the list. `EvtMax` is the max number of generated events which is different from the real number of generated events. Key4hep will try to simulate as much events with `EvtMax` as the limit, for high values of `EvtMax`, the number generated events will hit a ceiling that cannot be increased since Key4hep doesn't offer a way to set seeds.

A.4 Gaudi Consumer Algorithm

The Gaudi Consumer Algorithm whose purpose is to do the data selection for each generated event. To do the data selection, the `EDM4hep Reconstructed Particle` class was used (33). A full list of the variables produced are provided in the table 24.

Table 24 – Full list of variables produced by `MCConsumer` in the data analysis grouped by category.

Category	Variable	Description
Event counters	`m_nGenerated`	N_{gen}
	`m_nSelected`	N_{sel}
Global event	`nCharged`	N_{ch}
	`nChargedNonMuon`	N_{ch} (non-μ)
	`nTracks`	Reconstructed tracks
	`nTracksNonMuon`	Tracks (non-μ)
	`totalCharge`	Total charge
Visible system	`visibleEnergy`	E_{vis}
	`visiblePx`	p_x^{vis}
	`visiblePy`	p_y^{vis}
	`visiblePz`	p_z^{vis}
	`visiblePt`	p_T^{vis}
Observables	`m_invMass`	$m_{\mu\mu}$
	`m_recoilMass`	m_{recoil}
	`m_totalEnergy`	E_{tot}
Leading Muon	`m_mu1_E`	E^{lead}
	`m_mu1_px`	p_x^{lead}
	`m_mu1_py`	p_y^{lead}
	`m_mu1_pz`	p_z^{lead}
	`m_mu1_pt`	p_T^{lead}
	`m_mu1_charge`	q^{lead}
Subleading Muon	`m_mu2_E`	E^{sublead}
	`m_mu2_px`	p_x^{sublead}
	`m_mu2_py`	p_y^{sublead}
	`m_mu2_pz`	p_z^{sublead}
	`m_mu2_pt`	p_T^{sublead}
	`m_mu2_charge`	q^{sublead}
Muon pair	`m_px`	$p_x^{\mu\mu}$
	`m_py`	$p_y^{\mu\mu}$
	`m_pz`	$p_z^{\mu\mu}$
	`m_transverse_momentum`	$p_T^{\mu\mu}$
Missing	`m_missingEnergy`	E_{miss}
	`m_missingPx`	p_x^{miss}
	`m_missingPy`	p_y^{miss}
	`m_missingPz`	p_z^{miss}
	`m_met`	E_T^{miss}

A.4.1 Muon pair selection

The only criteria for the `MCConsumer` algorithm to decide which event was saved in the `ROOT` file was the presence of $\mu^+\mu^-$:

```
    if (mu1.charge * mu2.charge >= 0) {
    // same-sign or undefined charge
    return StatusCode::SUCCESS;
    }
```

this means that every event that didn't fit this criteria only contributed to the N_{gen} value. For events that did fit into the criteria, the invariant and recoil mass were calculate inside the algorithm:

```
    ...
    // Total 4-momentum of the muon system
    TLorentzVector dimuon = p4_1 + p4_2;
    m_invMass = (p4_1 + p4_2).M();
    ...
    // Initial 4-momentum of e+e- system
    TLorentzVector initial(0, 0, 0, m_ecm);
    // Recoil 4-momentum
    TLorentzVector recoil = initial - dimuon;
    m_recoilMass = recoil.M();
    ...
```

the `M()` function has functionality of the calculating the squared root of the module of any vector, resulting in the equations 4.22 and 4.33.

A.5 Run Time

The biggest drawback in Key4hep is a lackluster documentation, essential features such as Gaudi algorithms and `Collections` are easy to integrated, however, their purpose and where to find the available `Collections` and how to write a Gaudi algorithms was not clear.

The integration of `slcio` files into a form that it's readable in `EDM4hep` was the biggest pitfalls, it was necessary to introduce in the steering file the use of legacy features like `MarlinProcessors`, `Output_DST` and `LCIOWriter`, both which only had the purpose of serving as and intermediary to translate `slcio` into a readable form for the `MCConsumer` Gaudi algorithm, after that they produced two `slcio` files with an average size of $\approx 140\,\text{MB}$ with no utility for the data analysis. Lastly, after each process was simulated, a bash script `init.sh` would transfer the `hist.root` output into the respective output process folder and deleting the `test.slcio` and `output.slcio` files produced by `MarlinProcessors` to

free space; writing, saving and deleting this files was the most time consuming part of the simulation.

```
    # --- Marlin processing ---
    myProc = MarlinProcessorWrapper("Output_DST")
    myProc.ProcessorType = "LCIOOutputProcessor"
    myProc.Parameters = {
         "LCIOOutputFile": ["test.slcio"],
         "LCIOWriteMode": ["WRITE_NEW"]
    }
    algs.append(myProc)

    # --- LCIO output ---
    lcio_writer = io.add_lcio_writer("LCIOWriter")
    lcio_writer.Parameters = {
        "LCIOOutputFile": ["output.slcio"],
        "LCIOWriteMode": ["WRITE_NEW"]
    }
```

Listing A.6 – Key4hep section for the Marlin Processors.

For a full simulation run, done for all chosen backgrounds from Table 28 and signals, took on average 7 hours for a total of 124 processes.

Annex

ANNEX A – Particle Data Tables

Table 25 – Branching ratios relevant for invisible Higgs decay searches. Values are taken from the PDG 2024-2025 summary (14).

Particle	Decay mode	Branching ratio
Higgs (H)	$H \to \text{inv}$	$< 10.7\%$ (95% CL)
Z boson (Z)	$Z \to \mu^+\mu^-$	$(3.3662 \pm 0.0066)\%$

Table 26 – Cross sections for signal and background processes at $\sqrt{s} = 250\,\text{GeV}$. Electron beams are 80% polarized and positron beams are 30% polarized. Initial state radiation (ISR) is included, while *beamstrahlung* is not. Results obtained with WHIZARD 2.6.4 (2).

Process	σ_{LR} [fb]	σ_{RL} [fb]
$e^+e^- \to W^+W^-$	37500	2580
$e^+e^- \to e^+\nu W^-$	10200	1090
$e^+e^- \to e^+e^-Z$	3170	2000
$e^+e^- \to ZZ$	1800	827
$e^+e^- \to \nu\bar{\nu}Z$	220	13
$e^+e^- \to Zh$	313	211

ANNEX B – 250-SetA Folders Structure

The DESY ecosystem has the `250-SetA` folder is located at:
`/pnfs/desy.de/ilc/prod/ilc/mc-2020/ild/dst-merged/250-SetA/`
Inside the folder the child folders are divided into categories and within these folder the `slcio` files are found, table 27 has the folders inside `250-SetA`:

Table 27 – Folder structure for `250-SetA` grouped by fermion multiplicity in the final state.

2 fermions (2f)	4 fermions (4f)	3f / 5f / 6f / Other
2f_eehiq_eL_pL	4f_singleW_leptonic	3f
2f_eehiq_eL_pR	4f_singleW_semileptonic	5f
2f_eehiq_eR_pL	4f_singleZee_leptonic	6f
2f_eehiq_eR_pR	4f_singleZee_semileptonic	flavortag
2f_hadronic_eL_pR	4f_singleZnunu_leptonic	higgs
2f_hadronic_eR_pL	4f_singleZnunu_semileptonic	higgs_excl
2f_leptonic_eL_pR	4f_singleZsingleWMix_leptonic	
2f_leptonic_eR_pL	4f_WW_hadronic	
2f_Z_hadronic	4f_WW_leptonic	
	4f_WW_semileptonic	
	4f_ZZ_hadronic	
	4f_ZZ_leptonic	
	4f_ZZ_semileptonic	
	4f_ZZnunu_leptonic	
	4f_ZZnunu_semileptonic	
	4f_ZZWWMix_hadronic	
	4f_ZZWWMix_leptonic	
	aa_2f_Z_hadronic	
	aa_2f_Z_leptonic	
	aa_4f	

ANNEX C – Process Table

Table 28 – Summary of simulated processes, production IDs, and cross sections used in this analysis. Values for "Process ID", "Process Type" and cross-section σ where found at (15), only rows with "Energy" = 250 and "Polarization" = `eL.pR` where chosen. The "Number of Files" column was calculated with counting the number of files located at `250-SetA` for each "Process ID".

Process ID	Process Type	σ [fb]	Number of Files
402001	`e1e1h`	$(1.7671 \pm 0.0023) \times 10^{1}$	66
402003	`e2e2h`	$(1.6971 \pm 0.0014) \times 10^{1}$	64
402005	`e3e3h`	$(1.6941 \pm 0.0021) \times 10^{1}$	65
402007	`n1n1h`	$(6.0351 \pm 0.0062) \times 10^{1}$	62
402009	`n23n23h`	$(6.7111 \pm 0.0063) \times 10^{1}$	62
402011	`qqh`	$(3.4303 \pm 0.0044) \times 10^{2}$	143
402106	`e1e1h_aa`	$(4.0114 \pm 0.0052) \times 10^{-2}$	4
402107	`e1e1h_az`	$(2.7037 \pm 0.0035) \times 10^{-2}$	6
402101	`e1e1h_bb`	$(1.0285 \pm 0.0013) \times 10^{1}$	8
402102	`e1e1h_cc`	$(5.1071 \pm 0.0066) \times 10^{-1}$	8
402703	`e1e1h_dd`	$(9.1892 \pm 0.0119) \times 10^{-4}$	8
402109	`e1e1h_e2e2`	$(3.8524 \pm 0.005) \times 10^{-3}$	4
402108	`e1e1h_e3e3`	1.108 ± 0.0014	14
402103	`e1e1h_gg`	1.4473 ± 0.0019	10
402701	`e1e1h_ss`	$(4.3472 \pm 0.0056) \times 10^{-3}$	7
402702	`e1e1h_uu`	$(2.1206 \pm 0.0027) \times 10^{-4}$	8
402104	`e1e1h_ww`	3.7817 ± 0.0049	34
402105	`e1e1h_zz`	$(4.6299 \pm 0.006) \times 10^{-1}$	35
402142	`e2e2h_aa`	$(3.8523 \pm 0.0031) \times 10^{-2}$	3
402143	`e2e2h_az`	$(2.5965 \pm 0.0021) \times 10^{-2}$	6
402137	`e2e2h_bb`	9.8769 ± 0.008	8
402138	`e2e2h_cc`	$(4.9045 \pm 0.004) \times 10^{-1}$	8
402715	`e2e2h_dd`	$(8.8247 \pm 0.0072) \times 10^{-4}$	7
402145	`e2e2h_e2e2`	$(3.6996 \pm 0.003) \times 10^{-3}$	3
402144	`e2e2h_e3e3`	1.0641 ± 0.0009	14
402139	`e2e2h_gg`	1.3899 ± 0.0011	10
402713	`e2e2h_ss`	$(4.1748 \pm 0.0034) \times 10^{-3}$	7

Continued on next page

Process ID	Process Type	σ [fb]	Number of Files
402714	e2e2h_uu	$(2.0365 \pm 0.0017) \times 10^{-4}$	7
402140	e2e2h_ww	3.6317 ± 0.003	33
402141	e2e2h_zz	$(4.4463 \pm 0.0036) \times 10^{-1}$	34
402160	e3e3h_aa	$(3.8455 \pm 0.0047) \times 10^{-2}$	4
402161	e3e3h_az	$(2.5919 \pm 0.0032) \times 10^{-2}$	6
402155	e3e3h_bb	9.8595 ± 0.012	8
402156	e3e3h_cc	$(4.8959 \pm 0.006) \times 10^{-1}$	8
402721	e3e3h_dd	$(8.8092 \pm 0.0107) \times 10^{-4}$	8
402163	e3e3h_e2e2	$(3.6931 \pm 0.0045) \times 10^{-3}$	4
402162	e3e3h_e3e3	1.0622 ± 0.0013	15
402157	e3e3h_gg	1.3874 ± 0.0017	10
402719	e3e3h_ss	$(4.1674 \pm 0.0051) \times 10^{-3}$	8
402720	e3e3h_uu	$(2.0329 \pm 0.0025) \times 10^{-4}$	8
402158	e3e3h_ww	3.6253 ± 0.0044	35
402159	e3e3h_zz	$(4.4385 \pm 0.0054) \times 10^{-1}$	36
402178	n1n1h_aa	$(1.37 \pm 0.0014) \times 10^{-1}$	3
402179	n1n1h_az	$(9.2337 \pm 0.0095) \times 10^{-2}$	6
402173	n1n1h_bb	$(3.5124 \pm 0.0036) \times 10^{1}$	8
402174	n1n1h_cc	1.7441 ± 0.0018	7
402727	n1n1h_dd	$(3.1382 \pm 0.0032) \times 10^{-3}$	7
402181	n1n1h_e2e2	$(1.3156 \pm 0.0014) \times 10^{-2}$	3
402180	n1n1h_e3e3	3.784 ± 0.0039	12
402175	n1n1h_gg	4.9427 ± 0.0051	10
402725	n1n1h_ss	$(1.4846 \pm 0.0015) \times 10^{-2}$	7
402726	n1n1h_uu	$(7.2421 \pm 0.0074) \times 10^{-4}$	7
402176	n1n1h_ww	$(1.2915 \pm 0.0013) \times 10^{1}$	32
402177	n1n1h_zz	1.5812 ± 0.0016	33
402196	n23n23h_aa	$(1.5234 \pm 0.0014) \times 10^{-1}$	3
402197	n23n23h_az	$(1.0268 \pm 0.001) \times 10^{-1}$	6
402191	n23n23h_bb	$(3.9058 \pm 0.0037) \times 10^{1}$	8
402192	n23n23h_cc	1.9395 ± 0.0018	7
402733	n23n23h_dd	$(3.4898 \pm 0.0033) \times 10^{-3}$	7
402199	n23n23h_e2e2	$(1.463 \pm 0.0014) \times 10^{-2}$	3
402198	n23n23h_e3e3	4.2078 ± 0.004	12
402193	n23n23h_gg	5.4964 ± 0.0052	10
402731	n23n23h_ss	$(1.6509 \pm 0.0016) \times 10^{-2}$	7
402732	n23n23h_uu	$(8.0533 \pm 0.0076) \times 10^{-4}$	7

Continued on next page

Process ID	Process Type	σ [fb]	Number of Files
402194	n23n23h_ww	$(1.4362 \pm 0.0014) \times 10^{1}$	32
402195	n23n23h_zz	1.7583 ± 0.0017	33
402214	qqh_aa	$(7.7868 \pm 0.0099) \times 10^{-1}$	8
402215	qqh_az	$(5.2484 \pm 0.0067) \times 10^{-1}$	11
402209	qqh_bb	$(1.9964 \pm 0.0025) \times 10^{2}$	25
402210	qqh_cc	9.9136 ± 0.0126	15
402739	qqh_dd	$(1.7838 \pm 0.0023) \times 10^{-2}$	12
402217	qqh_e2e2	$(7.4781 \pm 0.0095) \times 10^{-2}$	8
402216	qqh_e3e3	$(2.1508 \pm 0.0027) \times 10^{1}$	35
402211	qqh_gg	$(2.8094 \pm 0.0036) \times 10^{1}$	17
402737	qqh_ss	$(8.4385 \pm 0.0108) \times 10^{-2}$	12
402738	qqh_uu	$(4.1164 \pm 0.0052) \times 10^{-3}$	12
402212	qqh_ww	$(7.3408 \pm 0.0094) \times 10^{1}$	59
402213	qqh_zz	8.9874 ± 0.0114	57
500002	2f_z_eehiq	$(1.9311 \pm 0.0011) \times 10^{5}$	4158
500010	2f_z_h	$(1.2797 \pm 0.0003) \times 10^{5}$	42130
500006	2f_z_l	$(2.1214 \pm 0.0012) \times 10^{4}$	2175
500118	4f_sw_l	$(3.422 \pm 0.0013) \times 10^{3}$	362
500106	4f_sw_sl	$(1.0264 \pm 0.0006) \times 10^{4}$	2737
500114	4f_sze_l	$(5.7747 \pm 0.0055) \times 10^{3}$	702
500102	4f_sze_sl	$(1.4233 \pm 0.0015) \times 10^{3}$	448
500126	4f_szeorsw_l	$(9.8431 \pm 0.0098) \times 10^{2}$	104
500122	4f_sznu_l	$(1.9513 \pm 0.0012) \times 10^{2}$	30
500110	4f_sznu_sl	$(4.5387 \pm 0.0026) \times 10^{2}$	143
500066	4f_ww_h	$(1.4866 \pm 0.0009) \times 10^{4}$	7390
500094	4f_ww_l	$(1.5634 \pm 0.0012) \times 10^{3}$	206
500082	4f_ww_sl	$(1.8779 \pm 0.001) \times 10^{4}$	5524
500062	4f_zz_h	$(1.4051 \pm 0.0008) \times 10^{3}$	810
500086	4f_zz_l	$(8.8957 \pm 0.0057) \times 10^{1}$	13
500074	4f_zz_sl	$(8.3808 \pm 0.0044) \times 10^{2}$	277
500090	4f_zznu_l	$(7.3399 \pm 0.0039) \times 10^{1}$	11
500078	4f_zznu_sl	$(6.0988 \pm 0.0039) \times 10^{2}$	192
500070	4f_zzorww_h	$(1.2389 \pm 0.0008) \times 10^{4}$	6179
500098	4f_zzorww_l	$(1.6371 \pm 0.0009) \times 10^{3}$	173
402309	6f_eexxxx	$(1.3531 \pm 0.0089) \times 10^{-2}$	3
402303	6f_eexyyx	$(7.1216 \pm 0.0407) \times 10^{-1}$	3
402315	6f_eeyyyy	$(1.5709 \pm 0.0103) \times 10^{-2}$	3

Continued on next page

Process ID	Process Type	σ [fb]	Number of Files
402307	6f_llxxxx	$(2.6534 \pm 0.0085) \times 10^{-2}$	3
402301	6f_llxyyx	1.1851 ± 0.0029	3
402313	6f_llyyyy	$(3.6334 \pm 0.0141) \times 10^{-2}$	3
402311	6f_vvxxxx	$(1.8949 \pm 0.0073) \times 10^{-2}$	3
402305	6f_vvxyyx	$(7.658 \pm 0.0321) \times 10^{-2}$	3
402317	6f_vvyyyy	$(1.1789 \pm 0.0051) \times 10^{-2}$	3
500318	2f-lcfi_z_bb	$(7.9514 \pm 0.0032) \times 10^{3}$	21
500314	2f-lcfi_z_cc	$(1.0759 \pm 0.0004) \times 10^{4}$	21
500302	2f-lcfi_z_dd	$(7.9514 \pm 0.0032) \times 10^{3}$	20
500310	2f-lcfi_z_ss	$(7.9514 \pm 0.0032) \times 10^{3}$	19
500306	2f-lcfi_z_uu	$(1.0759 \pm 0.0004) \times 10^{4}$	20
500338	4f-lcfi_zz_bbbb	$(7.0723 \pm 0.0061) \times 10^{1}$	23
500334	4f-lcfi_zz_cccc	$(5.6504 \pm 0.005) \times 10^{1}$	21
500322	4f-lcfi_zz_dddd	$(7.08 \pm 0.0062) \times 10^{1}$	20
500330	4f-lcfi_zz_ssss	$(7.08 \pm 0.0062) \times 10^{1}$	20
500326	4f-lcfi_zz_uuuu	$(5.6476 \pm 0.0049) \times 10^{1}$	20
410001	n23n23h_bb	$(3.9058 \pm 0.0037) \times 10^{1}$	55
410002	n23n23h_cc	1.9395 ± 0.0018	50
410005	n23n23h_dd	$(3.4898 \pm 0.0033) \times 10^{-3}$	48
410006	n23n23h_gg	5.4964 ± 0.0052	68
410003	n23n23h_ss	$(1.6509 \pm 0.0016) \times 10^{-2}$	47
410004	n23n23h_uu	$(8.0533 \pm 0.0076) \times 10^{-4}$	48

ANNEX D – Yields Evolution

The following tables document the full cutflow of the analysis, including both absolute weighted event yields and relative survival efficiencies for all considered process groups. They are provided to ensure full transparency and reproducibility of the event selection procedure.

Table 29 – Evolution of weighted event yields for signal and selected Higgs-related processes through the sequential application of selection cuts.

Requirement	Four-fermion (WW/ZZ)	Exclusive Higgs	Two-fermion
All Events	6.832×10^5	9.731×10^3	2.751×10^6
Muon Pair	6.832×10^5	9.731×10^3	2.751×10^6
Multiplicity N_{ch}	7.486×10^4	6.163×10^1	6.341×10^5
Muon $p_T > 10$	6.522×10^4	4.040×10^1	5.946×10^5
Missing energy	6.357×10^4	3.677×10^1	2.435×10^5
Transverse Momentum	4.565×10^4	2.662×10^1	4.266×10^3
Z mass window	6.150×10^3	8.188×10^0	4.612×10^2
Higgs recoil window	1.883×10^3	5.477×10^0	–

Table 30 – Evolution of weighted event yields for background processes through the sequential application of selection cuts.

Requirement	Four-fermion (single boson)	Other multi-fermion	Signal
All Events	3.285×10^5	1.151×10^2	2.553
Muon Pair	3.285×10^5	1.151×10^2	2.553
Multiplicity N_{ch}	5.105×10^4	–	2.496
Muon $p_T > 10$	5.817×10^3	–	2.372
Missing energy	2.472×10^3	–	2.372
Transverse Momentum	1.453×10^3	–	2.204
Z mass window	7.547×10^2	–	2.000
Higgs recoil window	1.908×10^2	–	1.665

Table 31 – Relative cut-by-cut survival efficiencies for signal and selected Higgs-related processes. Each entry corresponds to N_i/N_{i-1}.

Requirement	Four-fermion (WW/ZZ)	Exclusive Higgs	Two-fermion
All Events	1.000	1.000	1.000
Muon Pair	1.000	1.000	1.000
Multiplicity N_{ch}	0.110	0.006	0.231
Muon $p_T > 10$	0.871	0.656	0.938
Missing energy	0.975	0.910	0.410
Transverse Momentum	0.718	0.724	0.018
Z mass window	0.135	0.308	0.108
Higgs recoil window	0.306	0.669	–

Table 32 – Relative cut-by-cut survival efficiencies for background processes. Each entry corresponds to N_i/N_{i-1}.

Requirement	Four-fermion (single boson)	Other multi-fermion	Signal
All Events	1.000	1.000	1.000
Muon Pair	1.000	1.000	1.000
Multiplicity N_{ch}	0.155	–	0.978
Muon $p_T > 10$	0.114	–	0.951
Missing energy	0.425	–	1.000
Transverse Momentum	0.588	–	0.929
Z mass window	0.519	–	0.907
Higgs recoil window	0.253	–	0.833