Particle Flow: From First Principles to Gaugino Property Determination at the ILC

Dissertation

zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von

Mădălina Chera

aus Bukarest, Rumänien

Hamburg
2018

Gutachter der Dissertation:	Prof. Dr. Caren Hagner Dr. Jenny List
Gutachter der Disputation:	Prof. Dr. Elisabetta Gallo Dr. Jenny List
Datum der Disputation:	26.03.2018
Vorsitzender des Prüfungsausschusses:	Prof. Dr. Robin Santra
Vorsitzender des Promotionsausschusses:	Prof. Dr. Wolfgang Hansen
Dekan der Fakultät für Mathematik Informatik und Naturwissenschaften:	Prof. Dr. Heinrich Graener

This work is dedicated to the memory of two of my heroes: *Sir Terry Pratchett* and *Professor Umberto Eco*. They taught me, among many things, that discovering and accepting one's own humanity and sense of self can be as intricate as a next-to-leading order calculation and as rewarding as discovering a new particle. For that, and for writing some of the best literature there is, they have my eternal gratitude.

Je laisse Sisyphe au bas de la montagne! On retrouve toujours son fardeau.[...] Chacun des grains de cette pierre, chaque éclat minéral de cette montagne pleine de nuit, á lui seul, forme un monde. La lutte elle-même vers les sommets suffit á remplir un cœur d'homme. Il faut imaginer Sisyphe heureux.

Albert Camus - Le Mythe de Sisyphe, 1942

I leave Sisyphus at the foot of the mountain! One always finds one's burden again. [...] Each atom of that stone, each mineral flake of that night filled mountain, in itself, forms a world. The struggle itself toward the heights is enough to fill one's heart. One must imagine Sisyphus happy.

Albert Camus - The Myth of Sisyphus, 1942

Abstract

The International Linear Collider (ILC) is a planned electron-positron collider, which will enable high precision measurements of Standard Model (SM) observables, as well as direct searches for new particles beyond the SM. *Particle Flow* (PFlow) is a key concept for reaching the required level of precision at the ILC; one of its main aspects is that the energy of each particle should be determined in the optimal detector sub-component. The International Large Detector (ILD) is one of the two detector concepts proposed at the ILC. It was designed according to the principles of PFlow and constitutes the experimental context of this thesis.

The precision of the PFlow approach is limited by the *confusion* that can occur when reconstructing high energy events with a large particle multiplicity.

Detector simulations are used to do physics- as well as detector optimisation studies. While the full simulation presents a comprehensive description of the detector with a high level of realism, it is computationally expensive. Studies promptly requiring a large number of simulated events are only feasible by using a fast simulation, which uses a slightly simplified detector description. `SGV` is the fast detector simulation chosen for ILD. By default, it does not emulate the confusion effects. The impact of this simplification on the PFlow performance of `SGV` was investigated in this thesis. After the implementation of a Particle Flow confusion emulation routine in `SGV` ($\texttt{SGV}_{\texttt{PFL}}$), its performance was evaluated and compared to the full simulation.

With its default settings, the Particle Flow performance of `SGV` is up to 30% more optimistic than the full simulation. However, with the confusion implementation, the $\texttt{SGV}_{\texttt{PFL}}$ performance was on average 55% worse.

The Particle Flow performance of $\texttt{SGV}_{\texttt{PFL}}$ was further studied in the context of an NUHM supersymmetry scenario that considers the pair production of gauginos at the ILC. The $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ are assumed to be nearly mass degenerate and decaying into $W^{\pm}\tilde{\chi}_1^0$ and $Z^0\tilde{\chi}_1^0$, respectively. As a challenge for the Particle Flow reconstruction, the hadronic decays of the gauge bosons were chosen as signal. The performed analysis focused on determining the masses of the three gauginos and the cross-sections of the two signal processes. The same analysis methods were applied independently on Monte Carlo data produced both with the full simulation and with $\texttt{SGV}_{\texttt{PFL}}$. The resulting masses and cross-section values were compared. It was concluded that the large discrepancy between $\texttt{SGV}_{\texttt{PFL}}$ and the full simulation Particle Flow performance is not perceptible on an analysis level. The effect is obscured by much more prominent contributions from the jet clustering and jet-pairing effects.

Kurzzusammenfassung

Der International Linear Collider (ILC) ist ein geplanter Elektron-Positron Teilchenbeschleuniger, welcher Präzisionsmessungen von Observablen des Standardmodells (SM) sowie die Suche nach neuen Teilchen außerhalb des SM ermöglichen wird. *Particle Flow* (PFlow) ist ein Schlüsselkonzept um die benötigte Präzision am ILC zu erreichen; einer der Hauptpunkte ist, dass die Energie eines jeden Teilchens in der optimalen Detektorkomponente gemessen wird. Der International Large Detector (ILD) ist einer der zwei geplanten Detektorkonzepte am ILC. Der ILD wurde anhand des PFlow-Konzeptes entworfen und stellt den experimentellen Kontext dieser Arbeit dar.

Die Präzision des PFlow-Ansatzes wird durch *confusion* limitiert, die bei der Rekonstruktion von hoch energetischen Events mit großer Teichenmultiplizität auftreten kann.

Detektorsimulationen werden für Studien der Physik sowie zur Detektoroptimierung verwendet. Während die volle Simulation eine umfassende Beschreibung des Detektors mit hohem Realismusgrad liefert, ist sie sehr rechenintensiv. Studien, welche zeitnah eine große Zahl an simulierten Events benötigen, sind nur mit einer schnellen Simulation realisierbar, die eine vereinfachte Detektorbeschreibung beinhaltet. `SGV` ist die schnelle Simulation für den ILC. Standardmäßig emuliert sie keine confusion Effekte. Der Einfluss dieser Vereinfachung wurde in dieser Arbeit untersucht. Nach der Implementierung einer PFlow confusion Routine in `SGV` ($\texttt{SGV}_{\texttt{PFL}}$), wurde die Güte der Rekonstruktion mit der vollen Simulation verglichen.

Mit den Standardeinstellungen ist die PFlow Rekonstruktionsgüte um bis zu 30% übeschätzt, verglichen mit der vollen Simulation. Mit der confusion Implementierung ist die $\texttt{SGV}_{\texttt{PFL}}$ Rekonstruktionsgüte durchschnittlich 55% schlechter.

Weitere Untersuchungen der PFlow Rekonstruktionsgüte von $\texttt{SGV}_{\texttt{PFL}}$ wurden im Kontext eines NUHM Supersymmetrieszenarios, welches die Paarproduktion von Gauginos am ILC berücksichtigt. Die $\tilde{\chi}_1^\pm$ und $\tilde{\chi}_2^0$ werden als nahezu massenentartet angenommen und zerfallen in $W^\pm\tilde{\chi}_1^0$ und $Z^0\tilde{\chi}_1^0$. Als Herausforderung für die PFlow-Rekonstruktion wurden die hardronischen Zerfälle der Eichbosonen als Signal angenommen. Die Analyse befasste sich mit der Bestimmung der Massen der drei Gauginos und dem Wirkungsquerschnitt der zwei Signalprozesse. Dieselbe Analyse wurde unabhängig voneinander auf Monte Carlo Daten von der vollen Simulation und von $\texttt{SGV}_{\texttt{PFL}}$ angewandt. Die resultierenden Werte für Massen und Wirkungsquerschnitte wurden verglichen. Die Diskrepanz zwischen der vollen Simulation und $\texttt{SGV}_{\texttt{PFL}}$ ist nicht wahrnehmbar auf dem Level der Analyse. Sie wird durch stärkere Jetclustering und -pairing Effekte überlagert.

Contents

CHAPTER 1

Introduction

The *Standard Model* of particle physics provides the best description of our present understanding of the fundamental forces and elementary particles that form our Universe. The last prediction of the Standard Model to be confirmed experimentally was the existence of a particle that represented the most straightforward manifestation of the proposed Brout-Englert-Higgs mechanism.

The discovery of the Higgs boson at the CERN Large Hadron Collider in 2012 represented a significant turning point. With the addition of the Higgs boson, the full particle spectrum of the Standard Model is complete.

However, it has long been established that the description of nature offered by the Standard Model is not exhaustive. Presently, the supersymmetric extension of the Standard Model is among the favoured theoretical solutions that address the theory's shortcomings.

The Large Hadron Collider (LHC) is currently carrying out a wide range of searches for physics beyond the Standard Model, reaching for higher energies that could enable the direct production of new (supersymmetric) particles. At the proposed International Linear Collider (ILC), which forms the experimental context of this thesis, the nature of the electron-positron collisions enable different search strategies to be used.

At the ILC, polarised e^+e^- beams would collide at a tunable centre-of-mass energy spanning a range of 200-500 GeV and with the possibility of an upgrade to 1 TeV. In contrast to the LHC, where data analyses must handle challenging large and complex Standard Model backgrounds, the more benign collision environment at the ILC allows for a more straightforward event selection, with high purities and a background that is easier to compute.

The LHC and ILC experimental programs are complementary. Performing precision measurements at the ILC can not only provide evidence that particles with masses significantly higher than the e^+e^- collision energy exist, but it can also offer information concerning their properties.

Many interesting collision final states comprise hadronic jets. In order to reach the level of precision required at the ILC, the *Particle Flow* concept, which forms the main focus of this thesis, is crucial. First formulated in the CERN Large Electron Positron (LEP) era, it constitutes a different approach to measuring jet energies. The Particle Flow paradigm proposes that the precision with which the energies of hadronic jets are measured can be improved if the energy of each reconstructed particle in the event is measured in the detector subcomponent most suitable for that particle type. This has significant implications for the detector design.

The International Large Detector (ILD), i.e., the detector concept considered in this thesis, was specifically devised in view of the Particle Flow approach. The ILD design was optimised and its performance was evaluated in full simulation.

The simulation framework relevant for the work presented in this thesis consists of:

- Two subsequent versions of the ILD full simulation software. The most relevant differences between them are related to the degree of realism used in the detector descriptions.
- A fast detector simulation program called `SGV`. The capability to produce large amounts of Monte Carlo data in a short amount of time is particularly important since it enables the prompt investigation and update of the ILC physics potential, especially in view of the frequent actualisation of the results obtained by the LHC experiments.

The most mature software implementation of a Particle Flow reconstruction algorithm is PandoraPFA. In connection to the two full simulation versions mentioned above, two consecutive implementations of PandoraPFA were used.

The limiting factor of the Particle Flow approach is the phenomenon called *confusion*. In events with very high energies and particle multiplicities, the pattern recognition algorithms can no longer accurately distinguish between energy deposits coming from different particles. This leads to confusion errors like, e.g., the inadvertent splitting or merging of calorimeter showers, which can degrade the performance of the Particle Flow implementation.

In its default running mode, the fast simulation `SGV` does not consider these confusion effects. This simplification could lead to results that are too optimistic. The level of realism in the fast simulation must be increased such that its output becomes compatible with the Particle Flow performance of the ILD full simulation(s).

The first part of this thesis presents a series of studies carried out to: *(i)* investigate the impact of omitting the Particle Flow confusion in `SGV` on results obtained with the fast simulation, *(ii)* examine the feasibility of parametrising the confusion related behaviour of PandoraPFA with the aim of emulating it in the fast simulation and *(iii)* evaluate the Particle Flow performance of `SGV`, on first principles, after a confusion emulation routine was implemented.

The second part of the thesis focuses on the investigation of a supersymmetry (SUSY) scenario characterised by a final state that is particularly challenging for the Particle Flow reconstruction. The analysis was performed separately on

Monte Carlo data produced with `SGV` (with the confusion implementation) and with the two subsequent versions of the ILD full simulation. This is the first comprehensive comparison of the aforementioned three simulations. The goal was to evaluate the Particle Flow performance of the fast simulation in the context of a more involved physics study and to compare it with the output of the two full simulation versions. For this purpose, a significantly improved method for extracting the kinematic edges in fully hadronic decay chanels was used. The results obtained with the two full simulation versions were also compared.

CHAPTER 2

Theoretical Context

The theoretical foundations of this thesis are based on the *Standard Model* of particle physics which describes our current understanding of the fundamental building blocks of matter and the interactions between them. Since its first formulation [8–10], over fifty years ago, the theory has withstood numerous experimental tests, thus, becoming one of our most successful descriptions of nature. An overview of the Standard Model (SM) and its mathematical formulation is given in section 2.1 of this chapter.

Some of the Standard Model key aspects for the studies carried out and discussed in this thesis are the quark fragmentation and subsequent emergence of hadronic jets. These concepts are presented in section 2.2.

Despite its continuous confirmation through experimental results, it has become evident, over the years, that the description provided by the Standard Model is not complete. The theory's shortcomings are summarised in section 2.3. A potential solution to these issues postulates the existence of a new symmetry, i.e., *supersymmetry*. This extension to the Standard Model forms the context of the main physics study presented in this thesis. The most relevant theoretical aspects are summarised in section 2.4.

2.1 The Standard Model of Particle Physics

The elementary particles described by the Standard Model are considered to be point-like objects. Until now, there is no experimental indication either for their compositeness, i.e., the existence of an internal structure, or for their occupying a certain volume in space.

All elementary particles can be classified according to one of their intrinsic properties known as *spin*. The spin number is quantised and all particles have their spin expressed as a positive integer multiple of 1/2.

The particles that form matter are characterised by a spin of 1/2, i.e., an

odd multiple, equal to 1, and are called *fermions*. No elementary fermions with different spins, e.g., 3/2, 5/2, etc., are known to exist.

The fermions can be further categorised in *leptons* and *quarks*; the latter being the fundamental constituents of hadrons. Three particle generations exist for each category of fermions, as indicated in table 2.1. Each generation consists of a pair of particles that differ in terms of their quantum numbers. The corresponding particles from different generations have the same quantum numbers, but differ significantly in terms of their masses.

In addition, each fermion listed in table 2.1 has an anti-particle[1], i.e., a partner with the same mass but opposite quantum numbers.

Fermions	Generation	Particle Name	Mass	Charge	Spin
Leptons	I	e neutrino: ν_e	$<$2.2 eV	0	1/2
		electron: e	0.51 MeV	-1	1/2
	II	μ neutrino: ν_μ	$<$1.7 MeV	0	1/2
		muon: μ	105.66 MeV	-1	1/2
	III	τ neutrino: ν_τ	$<$15.5 MeV	0	1/2
		tau: τ	1.78 GeV	-1	1/2
Quarks	I	up: u	2.2 MeV	2/3	1/2
		down: d	4.7 MeV	-1/3	1/2
	II	charm: c	1.28 GeV	2/3	1/2
		strange: s	96 MeV	-1/3	1/2
	III	top: t	173.5 GeV	2/3	1/2
		bottom: b	4.18 GeV	-1/3	1/2

Table 2.1: Overview of the fermion content in the Standard Model. The mass values were taken from the summary tables of [11].

Fermions have another fundamental characteristic known as *chirality*. It is a Lorentz-invariant property that determines the manner in which the (quantum mechanical) wave function of a particle behaves when undergoing a transformation (rotation). There are two chirality eigenstates: a particle can be either *left-handed* or *right-handed*. In the massless particle approximation, the chirality eigenvalue coincides with the particle's helicity which is defined as the projection of the spin onto the direction of motion. Until now, all SM fermions have been observed in both helicity states with the exception of neutrinos which have been detected only in the left-handed state (while the anti-neutrinos are exclusively right-handed).

There are four fundamental forces acting on the matter forming particles mentioned above: gravity, electromagnetism, the weak force and the strong force.

On the spatial as well as on the mass scales relevant for particle physics, gravity is the weakest force and its effects are negligible. First formulated in its classical form in the seventeenth century, the most accurate description of gravity un-

[1] The discussion concerning the Dirac or Majorana nature of the neutrinos is beyond the scope of this overview.

til now was elaborated by A. Einstein in the theory of General Relativity. The success of the theory was experimentally demonstrated in 2015 when one of its main predictions, i.e., the existence of gravitational waves, was confirmed by the LIGO collaboration [12]. However, the formulation of General Relativity is very different from the quantum field theoretical apparatus of the Standard Model. Consequently, it was hitherto not possible to include a consistent and compatible description of gravity in the Standard Model.

Concerning the other three forces, their action is described as being mediated by particles called *bosons* which are characterised by an integer spin. The force carrying bosons couple to the fermions and are *exchanged* in the interactions between them. An overview of the force mediating bosons in the Standard Model is given in table 2.2.

Interaction	Particle Name	Mass	Charge	Spin
Weak	$W^{\pm}$	80.39 GeV	± 1	1
	Z^0	91.19 GeV	0	1
Electromagnetic	photon: γ	0	0	1
Strong	8 gluons: g	0	0	1

Table 2.2: Summary of the force mediating bosons in the Standard Model. The mass values were taken from [13].

The weak interaction has three mediators: two charged W bosons and the electrically neutral Z boson. They couple to all fermions with the addition that the $W^{\pm}$ bosons can also couple to each other.

The force carrier of the electromagnetic interaction is the photon which couples to all particles that have an electric charge, i.e., all fermions with the exception of neutrinos. The photon itself is massless and electrically neutral.

There are eight force mediators for the strong interaction called gluons. They couple exclusively to themselves and to quarks since these are the only particles in the Standard Model that carry the charge of the strong interaction known as *colour*. Unlike the photon, the gluons do carry the colour charge.

The theoretical descriptions of the electromagnetic and weak interactions were unified by S. L. Glashow [8], A. Salam [9] and S. Weinberg [10] into what is now known as the electroweak theory. Subsequently, the description of the strong force, formulated by H. Fritzsch, M. Gell-Mann and H.Leutwyler [14] in the theory of quantum chromo-dynamics (QCD), was added.

The basic elements of the mathematic formulation of the Standard Model are summarised in the following section.

2.1.1 Mathematical Formulation

The Standard Model of particle physics is a quantum field theory (QFT) in which the elementary particles introduced above are described as *quantum fields* defined at all points in space-time. Their dynamics and interactions are determined by the *Lagrangian density* (typically called the Lagrangian), $\mathcal{L}$, which acts similarly

to the way the Schrödinger equation does in non-relativistic quantum mechanics.

Symmetries

One of the fundamental concepts of the Standard Model is the principle of *local gauge invariance*. It was first described in the context of the electroweak theory. Therefore, in order to illustrate it, it is helpful to consider the Lagrangian of a free electron, i.e., a free spin 1/2 particle:

$$\mathcal{L}_D = i\bar{\psi}(x)\gamma^\mu\partial_\mu\psi(x) - m\bar{\psi}(x)\psi(x) \tag{2.1}$$

where $\psi(x)$ represents the electron field, expressed as a four-component Dirac spinor, γ^μ are the Dirac γ-matrices and m is the particle (electron) mass. $\mathcal{L}_D$ is invariant when applying a global gauge transformation, i.e., $\psi(x) \rightarrow \psi^{'}(x) = e^{i\alpha}\psi(x)$, considering a constant phase α. However, due to the derivative term in its expression, the free electron Lagrangian is not invariant under a local transformation: $\psi(x) \rightarrow \psi^{'}(x) = e^{i\alpha(x)}\psi(x)$, i.e., when the phase, α, is a function of space and time.

The invariance of $\mathcal{L}_D$ can be recovered by introducing a new field, A_μ, and replacing the standard derivative ∂_μ with the (gauge invariant) covariant derivative: $\partial_\mu \rightarrow D_\mu = \partial_\mu - ieA_\mu$. For this purpose, the introduced gauge field must transform as: $A_\mu(x) \rightarrow A^{'}_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu\alpha(x)$.

The newly obtained Lagrangian, $\mathcal{L}_{QED}$, is invariant under a local gauge transformation if the corresponding mass term for $A_\mu(x)$, i.e., $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$, is zero. Otherwise, the local gauge invariance would be once again broken. In fact, the introduced field corresponds to the photon which is indeed massless as demonstrated by the experimental results [11]. Lastly, it must be noted that the parameter e, i.e., the coupling strength, corresponds to the electron charge and it is a free parameter whose value must be determined experimentally.

The principle of local gauge invariance implies that the Lagrangian must be invariant under certain symmetry transformations. According to Noether's theorem [15], any fundamental symmetry corresponds to a conservation law. The mathematical description of symmetries in the Standard Model is achieved using group theory.

The symmetry group corresponding to the electroweak force is $SU(2)_L \otimes U(1)_Y$. In this context, $U(1)$ denotes a group of unitary 1×1 matrices and describes the weak hypercharge[2] conservation, i.e., only terms that are weak hypercharge neutral are allowed in the Lagrangian. The SU notation refers to a subgroup of U for which the condition $\det(U)=+1$ is fulfiled. The $SU(2)_L$ group describes the invariance under a left-handed weak isospin[3] transformation.

The gauge group underlying QCD is $SU(3)_C$. It describes the conservation of colour, i.e., the charge associated with the strong interaction. There are three possible colour charges, known as: *red*, *green* and *blue*.

[2]The *weak hypercharge* is a quantum number that relates the electric charge and the third component of the weak isospin: $Y_W = 2(Q - I_3)$.

[3]The *weak isospin* is a quantum number related to the weak interaction, analogous to the isospin concept in the context of the strong interaction. In case of the latter, the isospin classifies particles that are similarly affected by the strong force but have different charges.

Considering the conservation laws (symmetries) presented above, the gauge structure of the Standard Model can be formulated as:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{2.2}$$

The Standard Model Lagrangian

The Lagrangian formulation of the Standard Model ($\mathcal{L}_{SM}$) describes the dynamics of how the elementary particles, expressed as fields, evolve in spacetime and the manner in which they interact. Its complete mathematical description is highly involved and a detailed exposition exceeds the scope of this overview. The full expression of the Lagrangian can be found, for instance, in [16] or [17].

The Standard Model Lagrangian comprises: *(i) kinetic* and *self-coupling* terms that account for the "motion" of the fermion and gauge boson fields, *(ii) coupling* terms that describe how the gauge fields couple to the fermions, giving rise to interactions between them and *(iii) mass* terms which will be omitted in the current description and will be addressed in the next section.

In the following discussion, the relevant part of $\mathcal{L}_{SM}$ is the one containing the kinetic and self interaction terms. It can be expressed as:

$$\begin{aligned}
\mathcal{L}_{kin} &= \bar{\psi}\gamma^\mu D_\mu \psi - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \sum_{a=1}^{3} \frac{1}{4} W^a_{\mu\nu} W^{a,\mu\nu} - \sum_{a=1}^{8} \frac{1}{4} G^a_{\mu\nu} G^{a,\mu\nu} \quad \text{where:} \\
B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\
W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g \sum_{b,c=1}^{3} \epsilon^{abc} W^b_\mu W^c_\nu \\
G^a_{\mu\nu} &= \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_3 \sum_{b,c=1}^{8} f^{abc} G^b_\mu G^c_\nu
\end{aligned} \tag{2.3}$$

In the Standard Model, the fermion fields, ψ, account for the matter particles. In addition, there are four electroweak boson fields: W_1, W_2, W_3 and B. Furthermore, there are eight gluon fields, $G^{a=1\ldots8}$.

Thus, in equation 2.3:
→ The first term represents the kinetic term for a (Dirac) fermion.
→ The notation $B_{\mu\nu}$ denotes the gauge field tensor for $U(1)_Y$ (weak hypercharge) and B_μ is the gauge field. The coupling is typically denoted by g'.
→ The $W^a_{\mu\nu}$ notation represents the $SU(2)_L$ (weak isospin) gauge field tensor and W^a_μ is the gauge field. The coupling is indicated by g. The index ^{a} illustrates the three $W^{a=1,2,3}$ fields. The factor ϵ^{abc} represents the structure function of $SU(2)$
→ The gluon field tensor of $SU(3)_C$ (colour) is denoted as $G^a_{\mu\nu}$, where ^{a} runs over the eight colours and $G^{a=1\ldots8}_\mu$ represents the eight gluon fields. The coupling constant is denoted by g_3 and the structure function of $SU(3)$ is labeled as f^{abc} $(a, b, c = 1 \ldots 8)$.

For $G^{a=1\ldots8}_\mu$, there is a clear correspondence to the eight gluons (with eight different colour charges), introduced in table 2.2. However this is not the case for

the electroweak fields. In fact, they mix and thus create the physically observable states:

$$\begin{aligned} \gamma &= B\cos\theta_W + W_3\sin\theta_W \\ Z &= -B\sin\theta_W + W_3\cos\theta_W \\ W^{\pm} &= \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \end{aligned} \tag{2.4}$$

where γ, Z and $W^{\pm}$ denote the bosons introduced in table 2.2 and θ_W is the Weinberg (weak) mixing angle.

2.1.2 Electroweak Symmetry Breaking

The Standard Model Lagrangian, in its previously presented formulation, is gauge invariant only for massless fermions and gauge bosons. While the mediators of the electromagnetic (photon) and strong (gluons) interactions are indeed massless, this is not the case either for the fermions or for the three carriers of the weak force. Since their discovery in 1983 at the CERN SPS (Super Proton Synchrotron) by the UA1 [18], [19] and UA2 [20], [21] experiments, it is known that the $W^{\pm}$ and Z bosons have mass: $M_W = 80.385 \pm 0.015\,\text{GeV}$ and $M_Z = 91.1876 \pm 0.0021\,\text{GeV}$ [11], respectively. This experimental evidence is clearly in contradiction to the initial Standard Model prediction.

The theoretical solution for this discrepancy was formulated independently, in 1964, by P.W. Higgs [1], F. Englert and R. Brout [3] and G.S. Guralnik, C.R. Hagen and T.W.B. Kibble [22]. The proposed theoretical framework, typically known as the *Higgs mechanism*, enables the generation of the observed particle masses through spontaneous electroweak symmetry breaking.

In the Higgs mechanism, a complex $SU(2)$ doublet of scalar fields, as presented below, in equation 2.5, is introduced:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi_2 + i\Phi_1 \\ \Phi_4 + i\Phi_3 \end{pmatrix} \tag{2.5}$$

This is known as the *Higgs field* and it has four degrees of freedom illustrated by the real fields $\Phi_{i=1\ldots4}$. The corresponding Lagrangian is expressed as:

$$\mathcal{L}_{Higgs} = (\partial_\mu \Phi)^\dagger (\partial_\mu \Phi) - V(\Phi) \tag{2.6}$$

where the $V(\Phi)$ term represents the Higgs potential:

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{2.7}$$

In order to ensure that the potential is stable, λ, i.e., the Higgs self-coupling, must have a positive value.

If the μ^2 factor of the Higgs potential, shown in equation 2.7, has a negative value, the ground state, called the *vacuum expectation value*, is different from zero. In this case, the "choice" of a specific ground state, v, breaks the initial

potential symmetry in which $V(\Phi) = V(-\Phi)$. By minimising the potential, i.e., $\partial V/\partial\Phi = 0$, the vacuum expectation value can be obtained: $v = -\mu/\sqrt{\lambda}$. This implies that the Lagrangian from equation 2.6 is not gauge invariant and the $SU(2)_L \otimes U(1)_Y$ symmetry is broken. Furthermore, the spontaneous symmetry breaking of the Lagrangian generates a massless boson, known as the Goldstone boson, as follows from Goldstone's theorem [23].

P.W. Higgs [1], [2], F. Englert and R. Brout [3] demonstrated that the massless Goldstone boson must not necessarily manifest as a new (undiscovered) particle, but as a longitudinal degree of freedom of the gauge bosons, generating their masses.

One of the most important predictions emerging from the theory of the Higgs mechanism was the existence of a new scalar boson, named the Higgs boson, which was associated with the Higgs field.

In 2012, the prediction was successfully confirmed when the two general purpose experiments, ATLAS [4] and CMS [5], located at the Large Hadron Collider (LHC) at CERN, discovered a new particle that exhibited all the properties of the Standard Model predicted Higgs boson. Presently, a combined measurement of the Higgs boson mass, performed using $5\,\text{fb}^{-1}$ of data gathered at a centre-of-mass energy of $\sqrt{s}$ =7 TeV and an additional data set of $5\,\text{fb}^{-1}$ at $\sqrt{s}$ =8 TeV, from each experiment, revealed that the Higgs boson mass is $m_H = 125.09 \pm 0.21 \pm 0.11$ GeV [24].

2.2 Quark Hadronisation and Hadronic Jets

The existence of quarks and their classification scheme, known as the *"Eightfold Way"*, were proposed independently by M. Gell-Mann [25] and G. Zweig [26] in 1964. Quarks are the building blocks of strongly interacting particles, i.e., hadrons like, e.g., protons, neutrons, etc. The theory describing their interactions is called quantum chromo-dynamics (QCD) [14] and it represents a very important part of the Standard Model of particle physics.

The gauge group underlying QCD is $SU(3)_C$ and its dynamics is described by the fourth (last) term in the kinetic part of the Standard Model Lagrangian presented in equation 2.3.

The mediators of the strong force, the gluons, couple to each other and these self-interactions are believed to have an important effect on the behaviour of the QCD coupling. Thus, in comparison to the coupling of the electromagnetic force, for instance, the strength of the QCD coupling is small at high energy scales, i.e., at short distances (of the order of 10^{-17} m) between quarks. Consequently, the partons behave like quasi-free particles in these conditions. This phenomenon is known as *asymptotic freedom*.

In contrast, at lower energy scales, i.e., as the distance between the quarks increases, the strength of the coupling also increases significantly. This is closely related to the concept of *colour confinement* which denotes the observation that *all* hadrons in nature are colour singlet combinations of quarks and gluons. Even though the experimental evidence confirming the existence of quarks is indis-

putable, it is not possible to observe the presence of an individual (free) quark directly. Any attempt to separate or remove a quark from a colour-less state results in the immediate production of more (colour-neutral) hadrons. Presently, there is no univeral consensus concerning the theoretical explanation for the observed quark confinement. However, it is generally believed that it is related to the colour charge carrying nature of the gluons and to their ability to self-interact. An overview of the most prominent proposed explanantions can be found, for instance, in [27].

The processes that lead to the creation of new hadrons when partons are separated beyond the confinement distance are discussed below.

Quark Fragmentation

In order to illustrate the mechanisms underlying the quark fragmentation, it is useful to consider the process of e^+e^- annihilation to hadrons, depicted in figure 2.1.

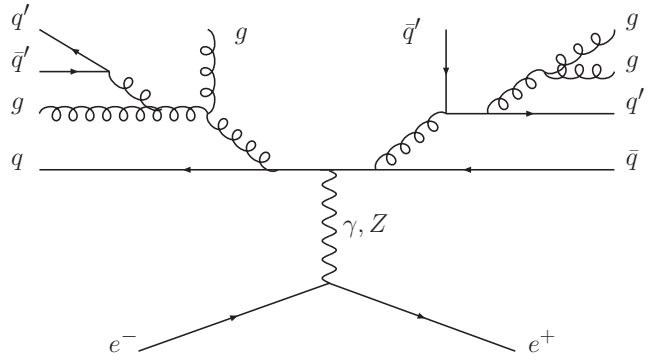


Figure 2.1: Sketch illustrating the quark-antiquark and gluon pairs cascades occuring during quark fragmentation. Figure taken from [28].

Initially, the two partons forming the quark-antiquark pair ($q\bar{q}$) are located at a very short distance from each other and, hence, behave as quasi-free particles. They have been highly accelerated in the production process and, as they decelerate, each of them radiates bremsstrahlung cascades of gluons (g) which are collimated into a cone characterised by a very small opening angle. As illustrated in figure 2.1, the emitted gluons can split further either into two gluons or in quark-antiquark pairs. This splitting process occurs repeatedly for as long as it is allowed by the amount of energy available.

At the same time, as the distance between the two initial quarks increases, the gluon field builds up a so-called *gluonic flux tube* between them. The flux tube has a very small transverse size and a high energy density of $\approx 1\,\mathrm{GeV}/10^{-15}\,\mathrm{m}$ [28]. When the separation distance between the two quarks reaches the value of $\approx 10^{-15}\,\mathrm{m}$ (i.e., the confinement radius) a sufficient amount of energy has accumulated in the gluon flux tube such that a new (complementary) quark-antiquark pair is spontaneously created and the tube "breaks", thus giving rise to two different hadrons. Consequently, the original quarks "fragment" into highly colimated cascades of quarks and gluons that will finally constitute hadrons. The emerging

hadronic jets preserve the momenta and energies of the initial quarks. Since the original individual quarks cannot be observed directly, their properties and/or kinematic characteristics can still be determined experimentally by measuring the properties of the associated hadronic jets. This consideration is one of the fundamental concepts underlying this thesis.

Hadronisation Models

The theoretical treatment of the correspondence between the initial quark and gluon configurations and the final, observable, hadronic jets requires non-perturbative mechanisms. Consequently, there is presently no rigorous theoretical (QCD) description of the manner in which hadrons are formed during the quark fragmentation. However, various hadronisation models that attempt to describe the involved phenomena as accurately as possible have been developed instead. The **independent jet fragmentation** model provides the gluon flux picture discussed above. In addition, there are two theoretical models widely used especially in Monte Carlo event generators:

→ **String Model**: The general approach implemented in the string model is similar to the one used by the independent jet fragmentation model (IJF). However, there are several significant differences. Firstly, the gluon flux tube described earlier is now pictured as a *string* between the $q\bar{q}$ pair. The energy dependent function describing the probability for the string to break also has a different expression. Furthermore, the gluons are treated differently: in the string model, they produce "kinks" which transfer energy and momentum to the strings, adding extra tension in them. The production of heavy quarks is suppressed while the creation of light quark pairs (u, d, s) is favoured. Lastly, the creation of baryons, i.e., hadrons containing three quarks, is performed by treating them as quark-diquark states where the diquark states are interpreted as antiquarks. A detailed description of the working principles of the string hadronisation model can be found, for instance, in [29].

→ **Cluster hadronisation**: The cluster hadronisation model relies on the concept of colour pre-confinement. Firstly, all radiated gluons are split into $q\bar{q}$ pairs until only quarks are present. The quarks are then clustered into colour-neutral groups with preferentially small invariant masses, although the distribution has a long tail towards higher mass values. The quark clusters then decay, usually producing two hadrons. The creation of heavy hadrons is generally suppressed by the cluster mass spectrum. Moreover, the heaviest clusters may first decay into smaller clusters and only afterwards produce the final hadron pairs. A detailed account of the cluster hadronisation model is given in [30].

2.3 Shortcomings of the Standard Model

The Standard Model of particle physics is a highly successful theory that provides an accurate description of many observed phenomena. Furthermore, its predictions have been confirmed by numerous experimental observations. The most recent validation of the Standard Model came from the observation and properties' measurements of the predicted Higgs boson [4], [5] in 2012. Nevertheless, despite its long lasting triumph, the theory does have several shortcomings, some of which are briefly summarised in this section.

→ **Grand Unification**: The successful merging of the theoretical frameworks describing the weak and the electromagnetic interactions led to the idea that a further unification with the theory of quantum chromo-dynamics could be possible at much higher energy scales. This would imply that, at high energies, all forces would couple with the same strength. However, as indicated in figure 6.8 of [31], the Standard Model equations that describe the evolution of the coupling constants at various energy scales show a clear divergence. Consequently, the Standard Model cannot accomodate the unification of all the three interactions. Furthermore, as mentioned earlier, the fourth fundamental force, gravity, also cannot be included.

→ **The Hierarchy Problem**: The Higgs boson mass receives large contributions from quantum loop corrections. Under the consideration that the Standard Model should be valid up to the Planck energy scale ($\Lambda_{\text{Planck}} \approx 10^{19}\,\text{GeV}$), these corrections become orders of magnitude larger than the experimentally determined mass of the Higgs boson, i.e., $m_H \approx 125\,\text{GeV}$. Reconciling the existing theory with the experimental observation involves an unnatural degree of fine-tuning that enables large cancellations to occur between tree level and higher loop orders, thus, keeping the Higgs mass at the right order of magnitude.

→ **Matter-Antimatter Asymmetry**: The Standard Model assumes that approximately equal amounts of matter and antimatter were produced at the beginning of our Universe. However, instead of annihilating each other immediately after their production, matter (i.e., baryons) clearly dominates our visible Universe. A potential explanation for the asymmetry involves the violation of the CP-symmetry (charge conjugation parity symmetry) [32], [33]. Nevertheless, until now, the effects of CP violation have only been observed in the context of weak interactions in the quark sector. These effects are not large enough to fully account for the observed matter-antimatter asymmetry.

→ **Neutrino Masses**: In its present formulation, the Standard Model does not accommodate neutrino masses. However, the experimental observation of neutrino flavour oscillations implies that the neutrinos cannot be massless. Data obtained from oscillation, neutrinoless double beta decay and cosmological experiments can provide upper limits on the sum of the neutrino masses and their ordering [34].

Presently, the experimental evidence indicates that, while the neutrinos are not massles, their masses must be about six orders of magnitude smaller in comparison to the other SM fermions.

→ **Dark Matter Candidate**: Many cosmological and astrophysical experimental observations like, for instance, the rotational velocity of galaxies [35] or the measurements of the anisotropy in the cosmic microwave background [36], indicate the presence of some unknown and invisible matter in the Universe, called dark matter. In fact, the amount of "visible" matter accounts for only ≈4% of the energy density in our Universe. Considering that the dark matter consists of previously undiscovered particles, any potential dark matter candidate must be massive and interact weakly. In this context, the only Standard Model possible contenders, the neutrinos, are directly excluded due to their very small mass.

2.4 Supersymmetry as an Extension of SM

The unsolved issues of the Standard Model, summarised in the previous section, indicate that there must be new physics not yet described by the theory. One potential solution, that could address many of the above mentioned problems, was proposed as the theory of supersymmetry (SUSY) [31] which correlates every known SM particle to a predicted super-partner (also called "*sparticle*").

The mapping of particles onto their corresponding super-partners (and vice versa) is described by the SUSY generator Q which acts as: $Q|\text{fermion}\rangle = |\text{boson}\rangle$ and $Q|\text{boson}\rangle = |\text{fermion}\rangle$, respectively. Thus, all SM fermions would have bosonic super-partners with the same masses, couplings and quantum numbers, clearly, with the exception of the spin which would differ by 1/2. Analogously, all SM bosons would have fermionic super-partners.

The introduction of sparticles under the new symmetry provides a natural solution to the hierarchy problem. One very important aspect of supersymmetry is that the quantum corrections cancel pair-wise between partners and super-partners due to a minus sign associated with the latter's fermionic loops. Consequently, the quadratically divergent loop contribution (mostly from heavy quarks) to the Higgs mass are automatically cancelled out by the contribution from the super-partners.

Furthermore, SUSY can also provide a potential dark matter candidate. If the lightest supersymmetric particle is stable (i.e., does not decay) and neutral it can only interact weakly, thus, being a likely candidate for dark matter.

While various implementations of supersymmetry have been proposed, the formulation most relevant for this thesis is the so-called minimal supersymmetric extension of the Standard Model (MSSM) which is described below.

2.4.1 The Minimal Supersymmetric Standard Model

The supersymmetric extension of the Standard Model considered in this thesis, MSSM [37], [38], is called minimal, since it introduces the smallest number of new particles while still ensuring that the theoretical model is phenomenologically feasible.

The super-partners of the SM leptons and quarks are called sleptons and squarks, respectively. Analogously, the fermionic super-partners of the gluons are known as gluinos while the bino and the winos constitute the corresponding super-partners of the gauge bosons. In MSSM, two Higgs boson doublets, with different hypercharge (± 1), are required in order to: *(i)* generate masses for up-type and down-type quarks and *(ii)* cancel the anomalies that occur when introducing the super-partners of the Higgs, known as higgsinos.

In any SUSY theory, it is convenient to write every field and its corresponding superpartner together as a *superfield* (*supermultiplet*). Thus, in MSSM, the gluons and their gluino super-partners, together with the electroweak gauge bosons and their gaugino fermionic super-partners constitute the gauge (vector) supermultiplets. The matter supermultiplets comprise the three generations of left-handed and right-handed quarks and lepton fields, their corresponding squark and slepton fields and the respective antiparticle fields. The field content in MSSM is summarised in table 2.3.

Supermultiplets	S = 0	S = 1/2	S = 1	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
gluon, gluino	-	$\tilde{g}$	g	8	1	0
winos, W boson	-	$\tilde{W}^0$, $\tilde{W}^{\pm}$	W^0, $W^{\pm}$	1	3	0
bino, B boson	-	$\tilde{B}$	B	1	1	0
squarks, quarks	$(\tilde{u}_L\ \tilde{d}_L)$	$(u_L\ d_L)$	-	3	2	+1/6
	$\tilde{\bar{u}}_L \equiv \tilde{u}_R$	$\bar{u}_L \equiv u_R$	-	$\bar{3}$	1	-2/3
	$\tilde{\bar{d}}_L \equiv \tilde{d}_R$	$\bar{d}_L \equiv d_R$	-	$\bar{3}$	1	+1/3
sleptons, leptons	$(\tilde{\nu}_L\ \tilde{e}_L)$	$(\nu_L\ e_L)$	-	1	2	-1/2
	$\tilde{\bar{e}}_L \equiv \tilde{e}_R$	$\bar{e}_L \equiv e_R$	-	1	1	+1
Higgs, Higgsino	$(H_u^+\ H_u^0)$	$(\tilde{H}_u^+\ \tilde{H}_u^0)$	-	1	2	+1/2
	$(H_d^0\ H_d^-)$	$(\tilde{H}_d^0\ \tilde{H}_d^-)$	-	1	2	-1/2

Table 2.3: Superfields in the MSSM model and the corresponding quantum numbers. In the case of quarks and leptons, only the first generation is presented. The anti-particles are not shown.

The gauge eigenstates presented in table 2.3 are allowed to mix in MSSM and they can form mass eigenstates. For instance, for any fermion f, the two supersymmetric partners ($\tilde{f}_L$ and $\tilde{f}_R$) which are the scalar super-partners of the respective left-handed and right-handed fermion can mix.

Neutralinos and Charginos

The neutral gauginos ($\tilde{B}$, $\tilde{W}^0$) and higgsinos ($\tilde{H}_d^0$, $\tilde{H}_u^0$) mix and thus produce the mass eigenstates known as *neutralinos*. Their respective mass term in the Lagrangian is: $\mathcal{L}_{\tilde{\chi}_0} = (-1/2)(\psi^0)^T \mathcal{M}_N \psi^0 + \text{h. c.}$, where $\mathcal{M}_N$ represents the neutralino mass matrix shown below in equation 2.8, expressed in the $\psi^0 = (\tilde{B}, \tilde{W}^0, \tilde{H}_d^0, \tilde{H}_u^0)$ basis [39, 40].

$$\mathcal{M}_N = \begin{pmatrix} M_1 & 0 & -m_Z \sin\theta_W \cos\beta & m_Z \sin\theta_W \sin\beta \\ 0 & M_2 & m_Z \cos\theta_W \cos\beta & -m_Z \cos\theta_W \sin\beta \\ -m_Z \sin\theta_W \cos\beta & m_Z \cos\theta_W \cos\beta & 0 & -\mu \\ m_Z \sin\theta_W \sin\beta & -m_Z \cos\theta_W \sin\beta & -\mu & 0 \end{pmatrix} \tag{2.8}$$

where M_1 and M_2 represent the bino and wino mass parameters of the MSSM, respectively, m_Z is the Z boson mass and θ_W denotes the weak mixing angle. The $\tan\beta$ parameter stands for the ratio between the vacuum expectation values of the two Higgs doublets and μ is the higgsino mass parameter.

By diagonalising the mass matrix, $\mathcal{M}_N$, four mass eigenstates (neutralinos) are obtained: $\tilde{\chi}_1^0$, $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$, $\tilde{\chi}_4^0$, enumerated in increasing order of their masses.

The nature of the neutralinos depends on the field content: if $|M_1| < \mu$ and $|M_2| < \mu$, the $\tilde{\chi}_1^0$ and $\tilde{\chi}_2^0$ are gaugino-like, while $\tilde{\chi}_3^0$, $\tilde{\chi}_4^0$ are higgsino-like. Depending on further SUSY model parameter values, the lightest neutralino, i.e., $\tilde{\chi}_1^0$ becomes the lightest supersymmetric particle (LSP).

Analogously, the charged winos and higgsinos mix and form four *charginos*. The corresponding mass term in the Lagrangian is given by:

$$\mathcal{L}_{\tilde{\chi}^\pm} = -\frac{1}{2}(\psi^+, \psi^-) \begin{pmatrix} 0 & \mathcal{M}_C^T \\ \mathcal{M}_C & 0 \end{pmatrix} \begin{pmatrix} \psi^+ \\ \psi^- \end{pmatrix} + \text{h. c.} \tag{2.9}$$

where the mass matrix, $\mathcal{M}_C$, can be written in the ψ^+=(-$i\tilde{W}^+$, $\tilde{H}_u^+$) basis as:

$$\mathcal{M}_C = \begin{pmatrix} M_2 & \sqrt{2} m_W \cos\beta \\ \sqrt{2} m_W \sin\beta & \mu \end{pmatrix} \tag{2.10}$$

Here, m_W denotes the W boson mass and the other notations are the same as in the case of equation 2.8. The outcome of the $\mathcal{M}_C$ matrix diagonalisation consists of two mass eigenstates, called *charginos*: $\tilde{\chi}_1^\pm$, $\tilde{\chi}_2^\pm$, mentioned here in the increasing order of their masses.

Considering the case when $M_2 < \mu$, the lighter chargino, $\tilde{\chi}_1^\pm$, has a wino-like nature, while the second one is higgsino-like.

An overview of the gauge and mass eigenstates in the minimal supersymmetric extension of the Standard Model is presented in table 2.4.

R-Parity

A new quantum number known as R-parity is introduced in the MSSM and it is defined as:

$$P_R = (-1)^{3(B-L)+2S} \tag{2.11}$$

Names	Spin	Gauge Eigenstates	Mass Eigenstates
Higgs bosons	0	H_u^0, H_d^0, H_u^+, H_d^-	h^0, H^0, A^0, $H^\pm$
Squarks	0	$\tilde{u}_L$, $\tilde{u}_R$, $\tilde{d}_L$, $\tilde{d}_R$ $\tilde{s}_L$, $\tilde{s}_R$, $\tilde{c}_L$, $\tilde{c}_R$ $\tilde{t}_L$, $\tilde{t}_R$, $\tilde{b}_L$, $\tilde{b}_R$	$\tilde{u}_L$, $\tilde{u}_R$, $\tilde{d}_L$, $\tilde{d}_R$ $\tilde{s}_L$, $\tilde{s}_R$, $\tilde{c}_L$, $\tilde{c}_R$ $\tilde{t}_1$, $\tilde{t}_2$, $\tilde{b}_1$, $\tilde{b}_2$
Sleptons	0	$\tilde{e}_L$, $\tilde{e}_R$, $\tilde{\nu}_e$ $\tilde{\mu}_L$, $\tilde{\mu}_R$, $\tilde{\nu}_\mu$ $\tilde{\tau}_L$, $\tilde{\tau}_R$, $\tilde{\nu}_\tau$	$\tilde{e}_L$, $\tilde{e}_R$, $\tilde{\nu}_e$ $\tilde{\mu}_L$, $\tilde{\mu}_R$, $\tilde{\nu}_\mu$ $\tilde{\tau}_1$, $\tilde{\tau}_2$, $\tilde{\nu}_\tau$
Neutralinos	1/2	$\tilde{B}^0$, $\tilde{W}^0$, $\tilde{H}_u^0$, $\tilde{H}_d^0$	$\tilde{\chi}_1^0$, $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$, $\tilde{\chi}_4^0$
Charginos	1/2	$\tilde{W}^\pm$, $\tilde{H}_u^+$, $\tilde{H}_d^-$	$\tilde{\chi}_1^\pm$, $\tilde{\chi}_2^\pm$
Gluino	1/2	$\tilde{g}$	$\tilde{g}$

Table 2.4: Summary of the gauge and mass eigenstates of the MSSM.

where B is the baryon number[4], L represents the lepton number[5] and S is the spin. In contrast to the lepton and baryon quantum numbers which are additive (i.e., it is their sum that is conserved in interactions), the R-parity is a multiplicative quantum number. All Standard Model particles, including the five predicted Higgs bosons, are characterised by P_R=+1 while the sleptons, squarks, gauginos and higgsinos have P_R=-1.

If R-parity is conserved, this has important implications: *(i)* when the initial states have even R-parity, the SUSY super-partners can only be produced in pairs, *(ii)* the SUSY particles can decay only into odd numbers of lighter SUSY particles and their SM counterparts and *(iii)* a lightest supersymmtetric particle should exist and it should be stable because any mixing betweent the SUSY and SM particles would be forbidden by the R-parity invariance.

2.4.2 Supersymmetry Breaking

If SUSY was an exact symmetry, the SM particles and their predicted super-partners would have the exact same masses. This assumption is clearly contradicted by the experimental observations: in contrast to their SM counterparts which have already been discovered, no super-partners have been detected until now. This implies that supersymmetry must be a broken symmetry.

The consideration that SUSY is *softly broken* requires the addition of explicit symmetry breaking terms in the supersymmetric Standard Model Lagrangian. Thus, the effective Lagrangian can be written as the sum between the unbro-

[4]The baryon number (B) is a strictly conserved quantum number defined such that all quarks are characterised by $B = +\frac{1}{3}$ while the antiquarks have $B = -\frac{1}{3}$. All the other fermions and gauge bosons of the Standard Model have a baryon number equal to zero.

[5]Analogously, the lepton number (L) is a conserved quantum number associated with the SM fermions denoted as leptons, i.e., $e, \mu, \tau, \nu_e, \nu_\mu$ and ν_τ. All leptons have L=+1 while their corresponding antiparticles have L=-1.

ken MSSM Lagrangian plus an additional Lagrangian that contains the SUSY violating mass terms: $\mathcal{L} = \mathcal{L}_{MSSM} + \mathcal{L}_{soft}$.

The source of the spontaneous SUSY breaking is called *the hidden sector*, in contrast with the Standard Model which is denoted as *the visible sector*. Similarly to the Higgs mechanism, some *messengers* transmit the SUSY breaking from the hidden sector to the visible one by means of loop corrections.

A large variety of models aiming to describe how SUSY is spontaneously broken and how the effects of the symmetry breaking are transported to the visible sector have been proposed. Three of the most developed and studied models are: *(i)* **mSUGRA**, the minimal Supergravity model [41], [42] in which the supersymmetry breaking is mediated through gravitational interactions, *(ii)* **GMSB** [43], [44] which considers that the supersymmetry breaking is communicated by means of the Standard Model's gauge interactions and *(iii)* **AMSB**, the anomaly-mediated supersymmetry breaking [45], [46], which represents a special flavour of the gravity mediated SUSY breaking being mediated through the conformal anomaly.

In the course of this thesis, the role of the jet reconstruction performance in reconstructing neutralino and chargino events will be discussed, using an mSUGRA example.

CHAPTER 3

The International Linear Collider

The Standard Model (SM) of particle physics, introduced in the previous chapter, is a well-established theory describing the interactions between elementary particles. Its predictions have been confirmed experimentally on numerous occasions. Nevertheless, it is understood that the model must be incomplete.

The search for answers to the open questions of the Standard Model led to the design and realisation of large-scale, high-energy particle physics experiments. One example for such a large-scale, collaborative endeavour is the Large Hadron Collider (LHC) currently running at CERN, with its two multi-purpose experiments (detectors), ATLAS [47] and CMS [48].

The focus of this thesis is one of the most mature next-generation collider projects [49]: the International Linear Collider (ILC). The ILC is a planned electron-positron linear collider with a foreseen centre-of-mass energy range between 200 GeV-500 GeV. The possibility of an upgrade to $\sqrt{s}$=1 TeV is also foreseen. In the context of the excellent performance of the LHC, it is important to remark upon the physics advantages of a linear lepton collider.

The ILC project offers excellent complementarity to the LHC. The discovery of a Higgs-like particle by the ATLAS [4] and CMS [5] collaborations in 2012 offers a good example. It is known that the ability to investigate the properties of the newly found particle at a hadron collider has intrinsic limitations. Here, the ILC would be able to provide valuable information, by performing model-independent, high precision measurements [50].

Furthermore, taking into account the present limits set by the LHC [51], the ILC has the potential of being a discovery machine itself. This will be illustrated later, in an overview of the ILC searches for physics beyond the Standard Model (BSM).

In parallel to the ILC, another linear collider project, called the Compact Linear Collider (CLIC) [52], is currently developed mainly at CERN. The two projects differ significantly in terms of accelerator technology and centre-of-mass energy

range. Nevertheless, there are a lot of common elements in terms of physics goals and detector technology and development. Therefore, both ILC and CLIC are now part of the Linear Collider Collaboration (LCC) [53].

The first detailed description of the ILC technical design was elaborated by the linear collider community and presented in the *Reference Design Report* [54, 55]. Many technical aspects of the ILC design, for example, the values of the beam parameters, were adjusted as the research and development efforts continued. The most recent description, followed in this chapter, was formulated in the *Technical Design Report* [49, 56, 57].

The current chapter offers an overview of the ILC experimental environment. The advantages of performing physics studies at a linear lepton collider are discussed in section 3.1. A brief technical description of the ILC is given in 3.2, followed by an overview of the specific experimental features in subchapter 3.3. The most important physics goals of the ILC are summarised in section 3.4 while the ILC running scenarios presently taken into consideration are presented in section 3.5.

3.1 Experimental Advantages

Lepton colliders have played a fundamental role in the development of particle physics. For example, the direct discovery of the gluon was achieved in 1979 at the Positron-Electron Tandem Ring Accelerator (PETRA) [58], [59] at DESY, in Hamburg. Another example of particular relvance for this thesis is the precise measurement of the W and Z-boson properties at the Large Electron-Positron collider (LEP).

Furthermore, the successful interplay between lepton and hadron colliders has been illustrated, e.g., by the prediction and subsequent measurement of the top quark mass. Using precise measurements of the Z boson performed at LEP and at the Stanford Linear Collider (SLC), the top mass was predicted from calculations of higher-order electroweak processes [60]. The predicted mass range was validated in 1995 by the discovery of the top quark at the proton-antiproton collider Tevatron, at Fermilab [61], [62].

All the lepton collider projects cited above have a common design feature: they are all circular colliders. Their main characteristic is that the bunches of colliding particles travel on a circular trajectory, in opposite directions, thus creating the possibility for the bunches to be collided again and again, a large number of times. This high collision rate enables the collection of a large amount of experimental data. However, the drive to explore physics at ever higher centre-of-mass energies is confronted with the intrinsic limitations of the circular design [63].

3.1.1 Advantages of a Linear Collider

The centre-of-mass energy reach of a circular collider is limited by the emission of synchrotron radiation. This type of electromagnetic radiation is emitted when charged particles are subjected to an acceleration perpendicular to their velocity.

This is clearly the case in a circular accelerator.

The amount of lost energy depends on the nominal beam energy (E_{beam}), the radius of the circular beam trajectory (R) and the mass of the accelerated particle (m), as shown in the equation below:

$$\Delta E \propto \frac{E_{beam}^4}{Rm^4} \tag{3.1}$$

The dependency on the particle mass creates a significant distinction between lepton and hadron circular colliders. Since the mass of an electron is almost 2000 times smaller than the proton mass, the expected energy loss via synchrotron radiation is much higher at a circular lepton collider. To counteract this effect, either the beam energy or the accelerator radius must be considerably increased.

An alternative idea, already proposed in the 1960's [63], is to consider a linear accelerator instead. In this case, $R \to \infty$ and thus no synchrotron radiation is emitted. However, the downside is that the particle bunches can be accelerated and collided only once. Nevertheless, besides solving the issue of synchrotron radiation, a linear collider also offers the benefit of more flexibility in terms of the maximum centre-of-mass energy due to its extendability.

The achievable energy is essentially directly proportional to the length of the accelerator. This implies that the construction costs follow approximately the same dependency. In fact, a cost optimisation study [64] showed that, for centre-of-mass energies above 200 GeV, a linear collider is more cost-effective than a circular one.

3.1.2 Advantages of a Lepton Collider

There are several distinct advantages that lepton colliders, regardless of their shape, provide with respect to hadronic machines. They are summarised in the following:

Clean final states:

In the present LHC running scheme, the proton bunches collide every 25 ns. Approximately 40 proton-proton interactions occur, on average, during each bunch crossing. Hundreds of energetic particles, resulting from each individual pp collision, essentially pile-up on top of each other, making the event reconstruction a difficult task.

In contrast, at the ILC, the electron and positron bunches collide every 300 ns and only one e^+e^- interaction takes place during a bunch crossing. Thus, the particle multiplicity in the event is significantly smaller. The only pile-up background expected at the ILC comes from $\gamma\gamma$ interactions. However, their cross-section is six orders of magnitude smaller than the total pp cross-section at the LHC. This means that, on average, one photon-photon collision is expected to occur per e^+e^- bunch crossing. This manifests through the presence of a few extra hadrons in the final state.

Furthermore, lepton collisions are not affected by issues such as beam remnants and multi-parton interactions.

Detailed knowledge of initial states:

In the case of hadronic reactions, the initial state is not well defined due to the composite nature of the interacting particles. On the contrary, at a lepton collider, the collision takes place between two elementary, point-like particles. Therefore, their initial energy is well defined and can be considered equal to the beam energy. Consequently, the full kinematics of the event can be reconstructed. This alows the further improvement of the reconstruction by performing constrained fits on the final state objects.

Democratic production of final states:

At a hadron collider like the LHC, the total cross section includes both the dominant non-perturbative, soft QCD processes and the high-energy processes that produce heavy particles in the final state. The latter type of interactions is more relevant for physics studies. However, the proton is a composite particle with a steeply falling parton distribution function. Consequently, the cross section for producing heavy particles is much smaller than for light ones. This indicates a clear hierarchy of the final states, where the ones involving light quarks and gluons are favoured.

From an experimental point of view, the interesting events are quite difficult to distinguish from the significantly larger QCD background. In order to handle this issue, the LHC experiments require a complex trigger system that discards all irrelvant events before the data analysis can proceed.

At a lepton collider, the coupling of the photon (and Z boson) to all SM and BSM particles is of the same order of magnitude. Thus, the production rates for both cases are similar and they are only limited by the available phase-space.
Even though the production cross-sections are generally smaller than in hadron collisions, there is no significant hierarchy between SM and new physics final states. Furthermore, the BSM events are clearly distinguishable from typical SM reactions. Consequently, the ILC can be operated without any triggering system. Moreover, all final states of a decaying particle can be recorded and used for further analysis. This makes it possible to measure the absolute branching ratios or total widths which is not achievable at the LHC.

More precise cross-section calculations:

The cross-sections calculations for hadronic interactions rely entirely on QCD. Their precision is affected by theoretical uncertainties arising from the proton structure functions and non-perturbative QCD effects. Furthermore, the NLO QCD corrections to the cross-sections are of the order of $\mathcal{O}(30-50\%)$. Reducing the theoretical uncertainties below 10% requires calculations at an NNLO level and even further. Due to the high complexity of these computations, the desired

level of precision is not feasible at the moment.

In comparison, the ILC offers the advantage of colliding elementary particles, without any uncertainties from parton distribution functions. The particle production relies on electroweak processes. The first order corrections to the cross-sections are on the level of only a few percent. Moreover, the theoretical precision can be improved to a part-per-mil level, using the calculations already performed for LEP. Together with the high experimental precision, this enables the ILC to study new BSM physics, even beyond its direct kinematic reach. Since both the theoretical, but also the experimental uncertainties are very small, the ILC can actively search for deviations from the precise SM predicted cross-section values. These small deviations would provide a clear indication of BSM physics.

Polarised beams:

The point-like nature of the colliding electrons and positrons provides the ILC with another important advantage: the e^+ and e^- beams can be polarised. At energies above the Z boson mass, especially, the left and right-handed quarks and leptons behave differently. This is highly useful, especially for BSM physics searches, where the polarisation of the incoming beams actively contributes to suppressing unwanted backgrounds.

3.2 The ILC Design

The ILC is presently the most advanced lepton linear collider project, embodying all the advantages summarised above. Figure 3.1 presents a schematic view of the ILC baseline where the most important accelerator components are highlighted. They will be briefly described in the current section. A more detailed review of the present accelerator design and considered technologies is presented in [55] and [57].

Electron and positron sources

The electron source comprises a polarised laser directed at a strained gallium-arsenide (GaAs) photocathode, where electrons are emitted via the photoelectric effect. This requires that the laser wavelength must be 790 nm. The properties of the target material as well as the laser polarisation ensure that 85% longitudinally polarised electrons can be produced in this setup. This is already above the ILC design requirement of 80% electron polarisation. The produced electrons leave the source with a low energy of only 200 keV. Before reaching the damping rings, they pass through the buncher - which produces the typical ILC bunch structure - and a pre-accelerator which increases the energy of the e^- bunches to 5 GeV.

Producing polarised positrons is a more elaborate process that uses fully accelerated electrons, i.e., with an energy of 125-250 GeV. For this reason, the positron source is located after the main e^- linac, as shown in figure 3.1. The electrons are directed through a $\approx$150 m long helical undulator where they produce cir-

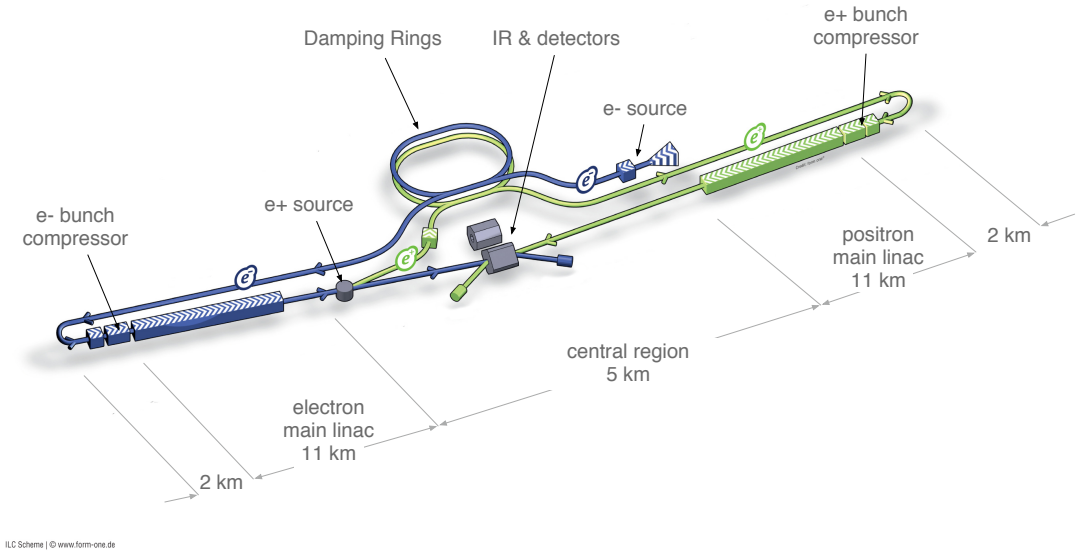


Figure 3.1: Sketch of the ILC layout. Picture taken from [49].

cularly polarised photons of 10-30 MeV via synchrotron radiation. The emitted photons then interact with a titanium-alloy target where they generate e^+e^- pairs. Only the positrons are retained, while the electrons and remaining photons are discarded in a beam dump. Similar to the e^- bunches, the positrons are pre-accelerated to 5 GeV before reaching the damping rings. This design of the positron source ensures a 30% longitudinal e^+ polarisation. Increasing the length of the helical undulator to ≈230 m would provide a positron polarisation of 60%.

At this stage, both the electron and positron beams have a high emittance. In order to bring this to the required level, they are guided towards the damping rings.

Damping rings

The damping rings are a very important component of any linear collider design. To reach the ILC luminosity goal, they must reduce the vertical emittance of the e^+ and e^- beams by five orders of magnitude, to the required 20 nm. The design comprises two damping rings, i.e., one for each beam, with a circumference of 3.2 km. Both rings are located in the same tunnel. The damping procedure consists of guiding each beam through a 100 m long wiggler. Traversing the wiggler causes the e^+ and e^- beams to emit synchrotron radiation with a much broader spectrum than in the case of an undulator. At the same time, a radio frequency (RF) module accelerates the beams only in their direction of motion. Consequently, the particle bunches become more parallel aligned and the emittance is reduced.

The electron and positron beams leave the damping rings with a much lower emittance, but with the same energy of 5 GeV. They are then transported over ≈15 km to the main accelerating systems.

Main Linacs

The low-emittance electron and positron bunches undergo another pre-acceleration stage (to 15 GeV) and length compression before reaching the main accelerators. Each main linac consists of approximately 7400 nine-cell niobium cavities. The cavities are each about 1 m long and they are cooled down to almost two Kelvin. At this very low temperature, the cavity material becomes superconducting which enables the use of a high electric field gradient of about 31.5 MeV/m. The niobium cavities are operated at an RF frequency of 1.3 GHz.

The European XFEL project [65], realised in collaboration with DESY, in Hamburg, uses the same accelerating technology as the ILC. This provides not only an excellent proof-of-principle, but also highly useful experience in constructing, comissioning and operating the accelerating facility.

Beam Delivery System

The final part of the accelerator is the beam delivery system (BDS). It is 3.5 km long and it first transports the electron and positron bunches from the main linacs to the interaction region (IR). The BDS incorporates several *beam diagnostics* tools:

(i) the **laser wire system** which can measure the beam size with a precision of 1 μm,

(ii) the **Compton polarimeter** which is able to perform beam polarisation measurements with an accuracy of $\Delta P/P \approx 0.25\%$,

(iii) the **energy spectrometer** which is required to measure the beam energy before (and after) the collision with a precision of $\Delta E/E \approx 10^{-4}$.

As they approach the interaction region, the e^+ and e^- beams traverse the final focussing system which comprises superconducting quadrupole and sextupole magnets. This brings the beams to the required size of a few nanometres. At the interaction point (IP), the electron and positron bunches collide under a crossing angle of 14 mrad, induced by a so-called crab cavity. This ensures the maximum overlap between the e^+ and e^- bunches and, consequently, the desired luminosity. The IR at the ILC is designed such that it allows the construction and operation of two general-purpose detectors. Each of them can be introduced or removed from the beam line via a push-pull system.

Downstream from the interaction point, a second polarimeter and beam energy spectrometer extract very useful information about what happened during the collision. Lastly, the highly disrupted electron and positron beams are discarded in a 11 m long, high pressure water beam dump. In its present design, the ILC beam dump can absorb up to 18 MW of power.

3.3 Experimental Features at the ILC

The experimental environment at a linear lepton collider has its own typical characteristics. Thus, the following section provides a brief overview of the standard experimental aspects at the ILC.

3.3.1 Beam Parameters and Luminosity

One of the most important physics goals of the ILC is to perform precision measurements. For this purpose, a high amount of a statistics, i.e., a large number of events, is required. For any given reaction, the number of events is directly proportional to the cross-section of the process and the luminosity: $N_{events} = \sigma \times \mathcal{L}$. Therefore, the luminosity must be as large as possible.

At a linear collider, the luminosity depends on several beam parameters, as shown in equation 3.2:

$$\mathcal{L} = \frac{n_b N^2 f_{rep}}{4\pi\sigma_x^*\sigma_y^*} H_D \tag{3.2}$$

Firstly, the beam structure plays an important role. In the typical ILC beam structure, a number n_b of electron (or positron) bunches, with each bunch containing N particles and having an average length σ_z, form a so-called *bunch train*. Within each bunch train, the separation between any two consecutive bunches is Δt_b. The bunch trains themselves are produced and accelerated at a repetition rate: f_{rep}.

Secondly, the luminosity also depends on the horizontal, σ_x^*, and vertical , σ_y^*, bunch sizes at the IP.

Lastly, the so-called enhancement factor, H_D, also determines the amount of luminosity. This factor accounts for the additional focussing of the e^+ and e^- beams, known as the *pinch effect*, and usually has a value of $H_D \approx 2$ [56]. The pinch effect is discussed in more detail in the following section.

The typical values of the ILC beam parameters mentioned above are summarised in table 3.1. Both the initial numbers, presented in the ILC *Reference Design Report* [54, 55], and the most recent ones, given in the *Reference Design Report* [49,56,57] are relevant for this thesis. The former were used in simulation-based physics studies for the ILD *Letter of Intent* (`LoI`) while the latter were similarly employed in the context of the *Detector Baseline Design* document (`DBD`). The quoted values in table 3.1 correspond to three different ILC operating scenarios, in the case of the `DBD`. Only the values corresponding to the 500 GeV centre-of-mass energy scenarios are relevant for this thesis.

It can be seen from equation 3.2 that, in order to raise the luminosity, one can: *(i)* increase the bunch train repetition rate f_{rep} as well the number of bunches per train n_b and *(ii)* strongly reduce the bunch size at the IP. The first option requires higher RF power and efficiency in converting it to beam power P_{beam}, where $P_{beam} = n_b N f_{rep} \cdot \sqrt{s}$. The latter option is reflected in the very small ($\mathcal{O}$(nm)) sizes that the ILC beam spots have at the IP. The extra focussing of

Design:			DBD			LoI
Centre-of-mass energy	$\sqrt{s}$	GeV	250	350	500	500
Design Luminosity	$\mathcal{L}$	$\times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$	0.75	1.0	1.8	2
Luminosity pulse repetition rate	f_{rep}	Hz	5	5	5	5
Bunch population	N	$\times 10^{10}$	2	2	2	2
Number of bunches	n_b		1312	1312	1312	2625
Bunch interval	Δt_b	ns	554	554	554	369
RMS horizontal beam size at IP	σ_x^*	nm	729	684	474	640
RMS vertical beam size at IP	σ_y^*	nm	7.7	5.9	5.9	5.7
RMS bunch length	σ_z	μm	300	300	300	300

Table 3.1: Overview of the ILC beam parameters. Both the original numbers, proposed in the RDR, and the most recent values, published in the TDR are quoted. The TDR numbers (denoted as DBD) were taken from [57] while the RDR values (denoted as LoI) were taken from [54].

the two incoming beams leads to a series of effects that are discussed in the next section.

3.3.2 Machine Backgrounds

The ILC provides a very clean experimental environment. Nevertheless, despite its considerably smaller multiplicities in the final state, the ILC is not a background free collider. The most important backgrounds stem from beam induced processes. The purpose of this section is to provide an overview of the most prominent backgrounds present at the ILC. A detailed description of all the machine backgrounds that can arise at the ILC is given in [57].

In order to achieve the design luminosity for the ILC machine, the colliding beams must be focussed to very small spot sizes. Consequently, this gives rise to a strong electrical field that accompanies the bunches. The approach of two highly focussed bunches disrupts the beam: the bunches exert an attractive force on each other which results in the individual particles inside one bunch to be accelerated towards the centre of the other in what is called the "*pinch effect*". Figure 3.2 shows a sketch illustrating this effect.

The occurence of the pinch effect does provide benefits: it helps reduce the bunch sizes even further which, in turn, leads to an increase in luminosity. This is expressed by the H_D factor in equation 3.2.

On the other hand, as the individual particles of a bunch are displaced by the electrical field of the other they emit photons, in what is essentially a special type of synchrotron radiation known as *beamstrahlung*. The photons emitted in this manner are focussed in a very forward direction and take away large amounts of energy. The average amount of energy lost by a particle via beamstrahlung is

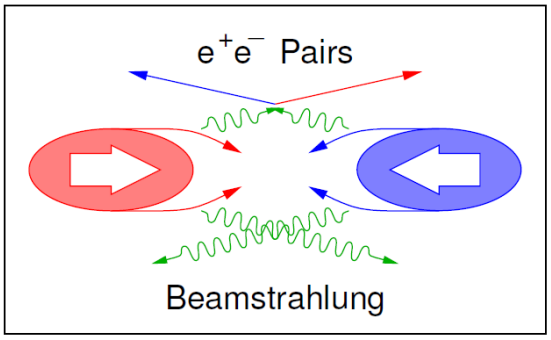


Figure 3.2: Sketch illustrating the pinch effect: both electrons and positrons are focussed at the centre of the bunch in the strong electrical field of the incoming bunch. Picture taken from [66].

given by the following expression:

$$\delta E_{loss} \propto \frac{\gamma}{E_{beam}\sigma_z}\left(\frac{N}{\sigma_x^* + \sigma_y^*}\right)^2 \tag{3.3}$$

where E_{beam} is the beam energy and the other variables are defined identically as above, in table 3.1.

On average, the energy loss of an electron (or positron) through beamstrahlung is in the order of $\approx 1\,\text{GeV}$ per particle. This leads to a decrease in the centre-of-mass energy that is available for the relevant physics interactions. From equations 3.2 and 3.3 it becomes apparent that the issue of energy loss via beamstrahlung can be overcome by making one bunch size larger with respect to the other. In the case of the ILC, the horizontal bunch size is much larger than the vertical size ($\sigma_x^* \gg \sigma_y^*$), as indicated in table 3.1.

The beamstrahlung photons are oriented in a very forward direction and many leave the interaction region through the beam-pipe. However, a non-negligible number do interact around the IP. These reactions produce two of the most pronounced backgrounds at the ILC: *(i)* the soft e^+e^- pairs and *(ii)* low transverse momentum $\gamma\gamma \rightarrow$ hadrons. They are briefly discussed in the following.

Electron-Positron Pairs

There are two distinct ways in which soft e^+e^- pairs can be produced in the vicinity of the ILC interaction point. In the first case, a beamstrahlung photon undergoes photon conversion in the electrical field of the electron or positron bunch. However, this process is negligible at the ILC [67]. The second production mode consists of two beamstrahlung photons scattering and thus creating the e^+e^- pair. In this case, there are three possible types of photon-photon interaction: *(i)* between two real photons (Breit-Wheeler process), *(ii)* between a real and a virtual photon (Bethe-Heitler process) and *(iii)* between two virtual photons (Landau-Lifshitz process). For the e^+e^- pair creation at the ILC, the last two processes are relevant while the real photon interactions contribute only at

the percent level [68]. Considering the typical ILC beam parameters, a number of approximately 10^5 e^+e^- pairs are expected to be produced per bunch. They carry an average energy of a few GeV per particle and are forward oriented, however not as much as the original beamstrahlung photons. Consequently, a fraction of the created e^+e^- pairs will be detected in the very forward components of the ILC detectors, constituting a source of background.

Low P_T Hadrons

The beamstrahlung photons can also interact and produce quark pairs that will subsequently hadronise. There are several processes through which the quark production can occur: *(i)* direct interaction, *(ii)* vector meson dominance and *(iii)* deep inelastic scattering. A detailed overview of these processes is provided in [69]. The $\gamma\gamma \rightarrow$ hadrons reaction has a large cross-section which increases with the ILC centre-of-mass energy. For the ILC centre-of-mass configuration relevant for this thesis, i.e., $\sqrt{s}$ =500 GeV, a number of $N_{\gamma\gamma}$=1.2 events are expected on average per bunch crossing. The outgoing hadrons typically have a low transverse momentum and have a forward focussed minijet topology. While their multiplicity is significantly lower than in the case of the e^+e^- pairs, the low P_T hadrons constitute an important background overlay to the physics events. A possible solution to removing the effects of the $\gamma\gamma \rightarrow$ hadrons background will be discussed in chapter 10 of this thesis.

3.4 Physics Goals of the ILC

The main three pillars of the ILC physics programme are: *(i)* precision measurements of the recently discovered Higgs boson, *(ii)* detailed measurements of the top quark properties and *(ii)* searches for new particles beyond the Standard Model (SM). The scope of the ILC physics goals is even broader, encompassing, among other topics, the precise study of W and Z bosons, detailed electroweak measurements, QCD tests, etc. A comprehensive account of the ILC physics outreach is given in [70] while an updated summary, from an experimental perspective, can be found in [50]. The phenomenological point of view is outlined in [71]. The following section provides an overview of the most important aspects of the ILC physics programme.

Higgs Physics

Measuring the properties of the recently discovered Higgs boson can conclude the electroweak symmetry breaking puzzle and provide new perspectives on physics beyond the Standard Model. The Higgs mass is the only parameter unconstrained by the SM, while all the other properties of the particle are predicted by the model. Presently, the mass of the Higgs boson has been measured by the LHC experiments with a precision of about 0.2% (M_H=125.09 GeV $\pm$0.24 GeV) [24]. Observing any deviations from the SM predictions, regarding the Higgs proper-

ties, would constitute strong evidence for physics beyond the Standard Model. This is where the ILC, with the advantages of a lepton collider, plays a very important role.

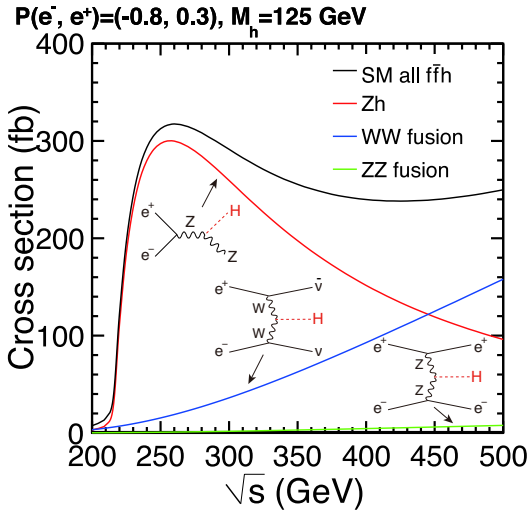


Figure 3.3: Cross-sections for the dominant Higgs production processes at the ILC as a function of the centre-of-mass energy. Plot taken from [50].

There are three main Higgs production modes at the ILC. Figure 3.3 shows their respective Feynman diagrams and cross-sections as functions of the ILC centre-of-mass energy. Thus, it can be seen that the *higgsstrahlung* process, i.e., $e^+e^- \rightarrow ZH$, is dominant at centre-of-mass energies lower than $\sqrt{s}$ =450 GeV while the other two reactions, i.e., $e^+e^- \rightarrow \nu\bar{\nu}H$, known as *WW fusion* and $e^+e^- \rightarrow e^+e^-H$ named *ZZ fusion*, become relevant at higher $\sqrt{s}$ values.

The higgsstrahung process enables the ILC experiments to measure the Higgs mass completely model independently. The mass determination is carried out by fitting the peak of the recoil mass distribution of the Z boson decay products. The nature of the Higgs decays is irrelevant to the analysis. Thus, no assumptions on the Higgs couplings or decay modes are necessary. Using this method, the Higgs mass can be measured with a precision of 30 MeV [50], i.e., approximately one order of magnitude better than the current LHC measurement. Furthermore, since all possible Higgs decay modes are considered, the invisible decay modes (e.g., involving dark matter candidate or other long lived particles) also contribute. Consequently, the higgsstrahlung process provides the opportunity to measure the inclusive $e^+e^- \rightarrow ZH$ cross-section without any model assumptions.

The total width of the Higgs boson, Γ_H, can be obtained by combining information from the measured production rates and branching ratios of several Higgs reactions, including higgsstrahlung and WW-fusion, observable at the ILC. The expeted precision on Γ_H is better than 2% [50].

The trilinear Higgs self-coupling is another very important measurement. Its value determines the shape of the potential energy of the Higgs field and consequently can provide information on the type of transition that occured in the

early Universe, leading to the current state of a broken electroweak symmetry. For this measurement, $e^+e^- \to ZHH$ is the relevant process. Considering the ILC running at 500 GeV centre-of-mass energy and combining only the $HH \to b\bar{b}b\bar{b}$ and $HH \to b\bar{b}WW$ processes, the current precision estimate for the trilinear Higgs self-coupling is 27% [72]. The precision improves at higher energies, i.e., $\sqrt{s}$ =1 TeV, where it can reach 10% [73] after a sufficient amount of data taking.

Top Quark Measurements

The top quark has a distinctive position in the Standard Model given by the fact that it is its heaviest elementary particle. Due to its high mass, the top quark couples most strongly to the Higgs boson. Consequently, it has the dominant contribution at loop level in the Higgs sector. Thus, the mass of the top quark plays a very important role not only in the SM, but also beyond. For example, precise measurements of both the Higgs and the top mass can shed light on the matter of the vacuum stability of the Universe [74].

At the ILC, the so called $1S$ mass of the top quark can be determined from studying the threshold-like behaviour of the top pair production cross-section with respect to the centre-of-mass energy, varied within a range of 10 GeV around $\sqrt{s}$=350 GeV. The $1S$ top mass can be converted to the well defined $\overline{MS}$ mass with an accuracy of about 10 MeV. The $1S$ mass itself can be measured with a precision of 50 MeV [75]. The study of the threshold region is illustrated by figure 3.4.

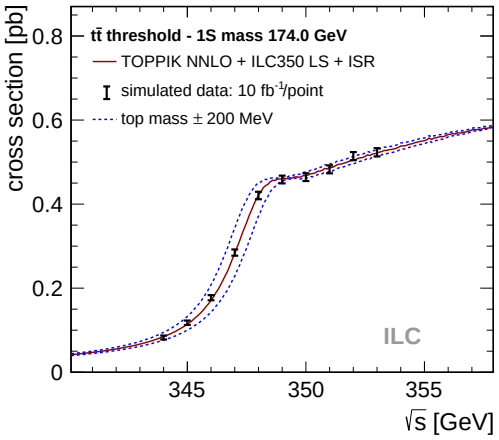


Figure 3.4: The threshold for top quark pair production at the ILC. Plot taken from [76].

This analysis cannot be performed at hadron colliders since tuning the centre-of-mass energy is not possible due to the composite nature of the colliding particles. It is estimated that at the High-Luminosity LHC (HL-LHC) the $\overline{MS}$ top mass can be extracted with a precision of 170 MeV from the invariant mass of its decay products in final states characterised by leptons and jets [77]. Thus,

the uncertainty on the top mass will be approximately three times higher at the HL-LHC than the expected ILC result.

Other key points of the ILC top physics programme consist of the measurement of the top Yukawa coupling and the top width Γ_t. The expected ILC precision is about 4.2% for the top Yukawa coupling and 26 MeV for Γ_t [50].

Physics Beyond the Standard Model

The ILC can access physics beyond the Standard Model indirectly, by performing very precise measurements of SM observables and searching for any potential deviations from their predicted values, but also directly by actively searching for new particles. If the BSM particles are within the kinematic reach of the ILC they can be precisely measured. The latter, direct approach is most relevant for this thesis. A detailed account of the ILC potential for discovering new physics is given in [70] and [78]. Two potential BSM scenarios are summarised in the following.

Weakly interacting massive particles (WIMPs) are a potential dark matter candidate. At a collider experiment, it is most likely that they would be produced in pairs. At the ILC, since the WIMP pair would leave no signal in the detector, such events could be identified when an initial state radiation (ISR) photon is also emitted. Thus, the relevant reaction is $e^+e^- \rightarrow \chi\chi\gamma$ and it would be observed as a single photon in an otherwise "empty" detector. Due to the clean experimental environment, in comparison to hadron colliders, such events would be observable at the ILC. The theoretical basis of this analysis relies on the so called "effective operators". The relevant parameter, Λ, can be interpreted as the energy scale at which the new physics occurs. It depends on the mediator mass ($M_{med.}$), the coupling to fermions (g_f) and the WIMP coupling (g_χ): $\Lambda = M_{med.}/\sqrt{g_f g_\chi}$. After accumulating 500 fb^{-1} at the ILC running at $\sqrt{s}$ =500 GeV, it is possible to reach an exclusion limit of Λ =2.5 TeV [79]. At a centre-of-mass energy twice as large ($\sqrt{s}$ =1 TeV), it is possible to exclude new physics up to Λ =4.5 TeV [79].

Searches for supersymmetric scenarios with highly compressed mass spectra are also part of the ILC BSM physics programme. For instance, a SUSY scenario in which the chargino and neutralino are light ($\mathcal{O}$=100 GeV), higgsino-like and almost mass degenerate was studied in simulation for the ILC running at a centre-of-mass energy of $\sqrt{s}$=500 GeV [80]. The analysis showed that the masses of the SUSY particles can be measured to a precision of 1.5-3.3 GeV and that the mass differences can be determined to 40-300 MeV. Thus, the advantages of a lepton linear collider make the discovery and identification of SUSY particles with only a few hundred MeV mass difference feasible. This is of particular importance for the study presented in this thesis.

3.5 ILC Operating Scenarios

Considering the rich ILC physics programme, it is important to establish several running scenarios which define how long the machine should be operated at a

certain centre-of-mass energy, which values of $\sqrt{s}$ are most relevant, which beam helicities would be used and how to share the amount of data taking time between the different energies and beam helicities. Several ILC operating scenarios have been studied by the ILC Parameters Joint Working Group and the conclusions have been presented in detail in [81].

It is essential to develop an operating scenario that optimises the ILC's physics potential. In view of this, the ILC Parameters Joint Working Group has recommended the scenario termed **H20**. This is the ILC running scenario most relevant for this thesis and it will be briefly summarised in the following.

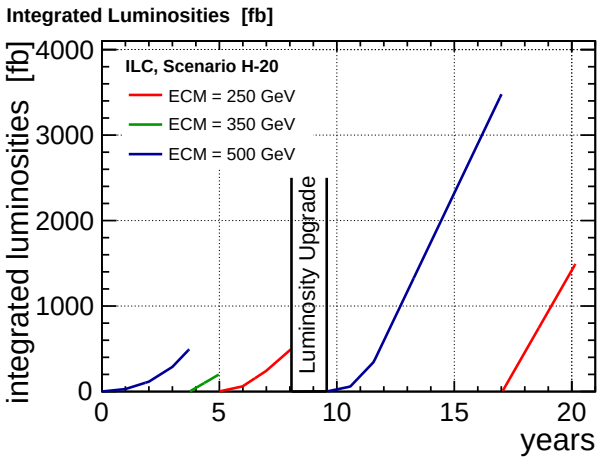


Figure 3.5: Graphic representation of the H20 ILC operating scenario showing the integrated luminosities to be collected at three different centre-of-mass energies and the corresponding necessary amount of time. Plot taken from [81].

A graphic summary of the H20 scenario is shown in figure 3.5. The initial phase consists of gathering 500 fb^{-1} of data at the centre-of-mass energy of 500 GeV. The sharing of the data taking time between different beam helicities is shown in table 3.2. This stage enables performing precise measurements of the Higgs coupling to the top quark, the Higgs self-coupling, but also direct searches for SUSY particles. The study presented in this thesis was performed in the context of this first phase.

In the next stage, 200 fb^{-1} of data will be recorded at the centre-of-mass energy of 350 GeV. This provides the opportunity to accurately measure the top quark mass by performing an energy threshold scan, but also to perform precision measurements of the W boson and Higgs couplings.

The third phase comprises the accumulation of 500 fb^{-1} at the centre-of-mass energy of 250 GeV which is the most relevant experimental context for detailed Higgs measurements.

Figure 3.5 also indicates the amount of time each stage is expected to take. Thus, the first three phases will be completed after approximately eight years. Following this, a period of ≈18 months will be dedicated to a complete shutdown and luminosity upgrade.

After the luminosity upgrade, further 3500 fb^{-1} will be collected at 500 GeV. This is particularly important for the precision of the top Yukawa coupling and

$\sqrt{s}[GeV]$		(-,+)	(+,-)	(-,-)	(+,+)
250	fraction [%]	67.5	22.5	5	5
	$\int \mathcal{L}dt$ [fb^{-1}]	1350	450	100	100
350	fraction [%]	67.5	22.5	5	5
	$\int \mathcal{L}dt$ [fb^{-1}]	135	45	10	10
500	fraction [%]	40	40	10	10
	$\int \mathcal{L}dt$ [fb^{-1}]	1600	1600	400	400

Table 3.2: The integrated luminosities that would be accumulated at three different centre-of-mass energies in the H20 running scenario. The fractions corresponding to different beam helicity configurations (P(e^-), P(e^-)) are also shown. Table adapted from [81].

Higgs self-coupling measurements. The last stage consists of gathering 3500 fb^{-1} with a centre-of-mass energy of 250 GeV. As indicated in figure 3.5, the H20 scenario will span approximately 20 years of the ILC operation.

The discovery at the LHC of a Higgs boson with a mass of approximately 125 GeV prompted the consideration that the construction of the ILC could be staged in order to reduce costs. For instance, the ILC project could begin as a "Higgs factory", initially built to achieve a centre-of-mass energy of 250 GeV, with the view of upgrading the energy reach up to 500 GeV or even 1 TeV at a later stage. A detailed description of the considered ILC staging scenarios is given in [82].

The advantages and physics capabilities of a linear lepton collider, like the ILC, have been illustrated throughout this chapter. However, in order to perform high precision measurements, the detector design and reconstruction techniques are also very important. Consequently, the concept of the "Particle Flow" reconstruction paradigm will be described next.

CHAPTER 4

The Particle Flow Paradigm

The physics programme of the International Linear Collider (ILC), presented in section 3.4, aims at providing insight and detailed information on a wide range of topics, e.g., fundamental Standard Model (SM) precision measurements, investigation of the electroweak symmetry breaking and searches for new physics beyond the SM (BSM).

One of the main advantages of the clean experimental environment and the well defined initial states at the ILC is the capability of inferring the existence of BSM physics if small deviations from the predicted values of SM observables are detected. This high level of precision can only be achieved if the properties of the particles emerging from the collision event, e.g., charge, position, momentum, energy, are measured very accurately.

The Particle Flow paradigm is one of the crucial factors that enable the required precision in measuring these event observables. The fundamental concepts of the Particle Flow philosophy constitute the focus of the present chapter.

The first part of the chapter presents several fundamental experimental characteristics of hadronic jets. The rationale that gave rise to the new approach is discussed next. The basic principles of the Particle Flow method are presented in section 4.3, while the requirements it places on the detector design are described in 4.4. Lastly, several of its past and present realisations at various experiments are discussed in section 4.5.

4.1 Experimental Aspects of Hadronic Jets

The fundamental building blocks of hadronic matter, i.e., quarks and anti-quarks, can only be observed experimentally as jets of particles. The theoretical framework concerning the quark (and gluon) fragmentation was discussed in section 2.2.

The link between the colourless, stable particles forming the jet and the original

partons is very important experimentally. For instance, highly relevant information regarding the kinematics of the initial partons and the potential presence of very short-lived BSM particles can be infered by reconstructing and measuring the jet.

Particle jets usually occur as a result of high energy hadron collisions when two partons (quarks *or* gluons), each a constituent of the two hadrons colliding, undergo a hard scattering process. They are also produced in e^+e^- annihilation processes, i.e., $e^+e^- \rightarrow q\overline{q}$. In both cases, as the partons accelerated in the scattering or production process start to decelerate, they give rise to highly collimated cascades of gluons and quark-antiquark pairs. At distances close to the colour confinement radius, i.e., $\approx 10^{-15}$ m, the partons produced in the cascade bind into hadrons which are also narrowly collimated.

The experimental context of producing jets of particles in hadron collisions is different from their production in e^+e^- annihilations. The main distinction is the existence of the so-called *underlying event* in the former case. Since hadrons are composite objects, the main hard scattering interaction is accompanied by the interactions of all the other *spectator* partons comprising the colliding hadrons. These interactions are essentially not correlated with the main hard scattering and constitute the underlying event.

The situation is different at a lepton collider like the planned ILC where, due to the point-like nature of the colliding particles, there is no underlying event. However, even in the relatively clean environment of a lepton collider, the final state manifestation of a hadronic jet can be affected by beam backgrounds like, e.g., $\gamma\gamma \rightarrow hadrons$, described in section 3.3.2.

From an experimental perspective, the presence of a jet appears in the form of a large number of tracks and calorimeter energy deposits concentrated in a region of the detector, as the particles that constitute the jet interact in the sensitive layers.

The first experimental evidence for the occurrence of hadronic jets was provided in 1975 by the e^+e^- collider SPEAR [83], [84], located at SLAC. The SLAC-LBL Collaboration used events observed by the MARK I detector [83] to demonstrate that the final state hadrons were not uniformly distributed in the detector but concentrated in the vicinity of the event axis which was determined by the momenta of the initial quarks. Furthermore, the charge analyses and angular distributions measured by the PLUTO Collaboration at the DORIS II e^+e^- accelerator running at DESY delivered crucial evidence [85] in 1978 that the observed jets of hadrons are indeed associated to the initial quarks.

4.1.1 Jet Definitions

Even though the presence of hadronic jets in the experimental data is qualitatively obvious, the jets are not well defined objects. While it is known that an individual quark or gluon can never be directly observed due to colour confinement, a collimated jet of stable particles is believed to be the unequivocal result of a single parton, temporarily isolated in momentum space, undergoing frag-

mentation. However, an individual coloured object like a quark or a gluon cannot fragment into colour-neutral hadrons in complete isolation. Such a process would not fulfil the laws of colour, energy and momentum conservation. Therefore, the fragmentation process must involve some form of participation from the other surrounding partons. Consequently, this introduces an unavoidable degree of ambiguity in defining a jet already on the theoretical level.

From the experimental point of view, the closest achievable practical definition of a jet is represented by the set of criteria implemented in the jet clustering algorithms. Their task is to group the stable final-state particles, observed and measured in the detector, into structures that would ideally resemble the original, true jets of particles as closely as possible. The jet algorithms follow a set of well motivated rules that determine how the particles should be grouped and how to assign momenta to the resulting jets. These rules represent the jet definition of the respective clustering algorithm.

The first endeavour to formulate a fundamental, standardised and quantitative definition of a jet materialised during the 1990 *Snowmass Summer Study on High Energy Physics* [86]. Over the years, motivated by the specifics of the running experiments, many jet definitions have been developed and implemented. Presently, the vast majority of jet clustering algorithms can be classified in two categories described below.

Cone algorithms define jets as the dominant directions in which the final state particles emerge after hadronisation. They use the concept of a *stable cone* defined as a circle in the η-ϕ plane, where ϕ is the *azimuthal angle* around the beam axis and η represents the *pseudorapidity*[1]. A cone is stable if the sum of all the momenta of the particles present in the cone points in the direction of the circle's centre. Based on this definition, the algorithms try to identify all the stable cones in the event. The possibility that two stable cones overlap is also accounted for. In most cases, in order to make the cone algorithms computationally fast, a seeding procedure is necessary.

While cone algorithms used to be favoured due to their straightforward implementation, they present a significant disadvantage since the majority are *infrared* and *collinear* unsafe. In the case of jet algorithms, the concept of infrared and collinear safety (IRC) is expressed as the requirement that the set of obtained final jets remains unchanged by the emission of an extra soft particle or by the collinear split of one of the particles, respectively. If the jet algorithm is IRC unsafe both the infrared emission and the collinear splitting lead to a different configuration of the final state jets. The perturbative QCD consequence of this is the lack of cancellation of infrared and collinear infinities, thus giving rise to divergent cross-sections. A more detailed discussion on the issue of IRC safety can be found in [87]. The experimental implications of using an IRC unsafe jet

[1] The pseudorapidity, η, is an observable that describes the angle of a particle with respect to the beam direction. It is defined as $\eta = -\ln\left[\tan\frac{\theta}{2}\right]$ where θ is the polar angle relative to the positive direction of the beam axis, in the detector coordinate system. For example, at θ=90°, the pseudorapidity is zero.

algorithm are that the jet measurements become impossible to compare with the theoretical calculations.

A solution was proposed in 2007 in the form of the Seedless Infrared-Safe Cone (SISCone) algorithm [88]. The SISCone approach avoids the dependence of the final state jets configuration on the use of stable particles as seeds and subsequent iterations which can lead to IRC unsafe results and aims at finding all stable cones instead. A detailed description of the SISCone algorithm and its performance can be found in references [87] and [88].

Sequential recombination algorithms are based on the assumption that the particles emitted in a jet will have rather similar transverse momenta (i.e., they travel in approximately the same direction). Therefore, the prescription according to which particles are clustered is formulated in terms of the transverse momentum.

All sequential clustering algorithms follow the same procedure. Two distance variables are defined. The first one represents the distance in momentum space between any two particles i and j: $d_{ij} = \min(p_{T_i}^a, p_{T_j}^a) \cdot \frac{R_{ij}^2}{R}$, where a is an exponent that varies according to the jet algorithm, for example, a has a value of two for the k_T algorithm [89] and zero for the Cambridge/Aachen algorithm [90]; R_{ij}^2 is the distance in the η-ϕ plane between the two particles and R is the radius parameter that determines the final size of the jet. The second variable, d_{iB}, is defined as the distance in the momentum space between the beam axis and particle i: $d_{iB} = p_{T_i}^a$.

The algorithm then proceeds to find the minimum between d_{ij} and d_{iB}, considering the list of all the particles observed in the event. If d_{ij} is the minimum then particles i and j are first combined into a *pseudojet* that replaces them. The initial i and j particles are then removed from the list. However, if d_{iB} is the minimum, particle i is classified as part of the beam jet and is then removed from the list of particles. The procedure is repeated either until all particles are assigned to a jet or until the required number of jets was found.

In contrast to most cone algorithms, the sequential recombination algorithms do not require any seeding and are IRC safe.

4.1.2 Jet Energy Resolution

Once the jet definition has been applied to the observed particles in an event, the properties of the obtained jets can be studied. The jet energy provides very useful information about the kinematics of the initial interaction. Thus, in order to determine the specifics of the collision event, the produced jets must be reconstructed and measured as accurately as possible.

However, the precision with which the jet energy can be measured is limited by several sources of uncertainty, presented below. The total imprecision in measuring the energy of a reconstructed jet is expressed as the jet energy resolution, $\sigma_{\mathrm{E_j}}$.

→ Parton Fragmentation and Hadronisation Uncertainties:
The quantum chromodynamics (QCD) mechanisms responsible for transforming

initial parton configurations into hadronic jets cannot be treated in a theoretical rigorous manner. To compensate this fact, several parton fragmentation and hadronisation models have been formulated with various degrees of sophistication, as discussed in section 2.2. These models have been implemented and used in Monte Carlo event generators. However, none of them can perfectly describe the occuring processes. These unavoidable discrepancies could lead to wrong predictions and an inaccurate outcome of the jet simulation. Since many analyses rely either exclusively or to a large extent on Monte Carlo jet simulations the modelling uncertainties ($\sigma_{\text{Modell.}}$) contribute to the total jet energy resolution.

→ Jet Clustering Errors:
Ideally, a perfectly clustered jet would comprise *all* the particles produced *exclusively* in the initial hadronisation process, without any contamination. However, from an experimental point of view, there is a high degree of variability, on a jet to jet basis, in terms of properties like, e.g., the hadron multiplicity, the spatial spread of the final state particles, etc. These fluctuations can affect the performance of the jet clustering. Further confusion is added by the fact that, in collision events, other hadrons produced in beam background interactions or from the underlying event can overlap with the jets. Similarly, particles from different jets can also overlap. Thus, these represent significant challenges to any jet clustering algorithm which can err both by accidentally adding particles that are not part of the jet or by wrongly excluding some that are. This clearly introduces further uncertainty ($\sigma_{\text{Clust.}}$) in the determination of the jet energy.

The ILC community is presently developing a software tool called `TrueJet` [91] that can be used on simulated data to link the Monte Carlo truth jet information, i.e., the true Monte Carlo particles forming the jet, to the output of the event reconstruction and subsequent jet clustering. This will enable a quantitative estimation of the impact the jet clustering errors have on the jet energy resolution.

→ Detector Resolution:
Traditionally, the energy of a jet is determined by adding up the amounts of energy deposited in the electromagnetic and/or hadronic calorimeter (ECAL and HCAL) by the particles identified as part of the jet. Thus, the jet energy determination relies almost entirely on the calorimetric measurements. Consequently, the jet energy resolution is significantly affected by the calorimeters' energy precision ($\sigma_{\text{Det.}}$) which can be parametrised as follows:

$$\frac{\sigma(E)}{E} = \frac{\alpha}{\sqrt{E}} \oplus \beta \oplus \frac{\gamma}{E} \tag{4.1}$$

The α factor in the above equation is the *stochastic term* which accounts for the intrinsic statistical fluctuations of the calorimetric shower, the sampling fluctuations, etc. The *constant term*, β, takes into account the calibration inhomogeneities as well as the leakage and energy loss effects in regions with dead material. Lastly, γ is the *noise term* which accounts for read-out electronics noise effects.

For illustration, some typical values of the calorimeter single particle energy resolution, obtained at previous and currently on-going experiments, are given in table 4.1.

Experiment	$\sigma(E)/E$ for ECAL			$\sigma(E)/E$ for HCAL		
	α_{ECAL}	β_{ECAL}	γ_{ECAL}	α_{HCAL}	β_{HCAL}	γ_{HCAL}
ALEPH	18%	0.9%	-	85%	-	-
ZEUS	18%	1%	-	35%	2.0%	-
H1	11%	0.6%	154 MeV	51%	1.6%	900 MeV
ATLAS	10%	0.4%	300 MeV	45%	1.3%	-
CMS	3.5%	0.66%	185 MeV	90%	4%	-

Table 4.1: Typical single particle energy resolutions achieved by the electromagnetic and hadronic calorimeters at the ALEPH [92], ZEUS [93], [94], H1 [95], [96], ATLAS [97] and CMS [98] experiments.

The values quoted in table 4.1 are usually obtained by studying the detector performance in test-beam campaigns which consist of aiming single particle beams, with a well known and tunable initial energy, directly at the calorimeter under study. The ECAL energy resolution is typically determined using electrons as in the case of ALEPH [92], for instance, while protons and charged pions are commonly used for evaluating the HCAL resolution.

→ Heavy Quark Decays
When heavy quarks are present after the hadronisation process, e.g., in the form of B mesons, they decay semileptonically with a probability of about 10% [99] to a charged lepton, the corresponding neutrino and a small number of hadrons. Since neutrinos cannot be detected nor measured, their presence in the final state affects the precision with which the jet energy can be measured. A basic solution to estimate and account for the missing neutrino energy, in the experimental context of the ILC, is proposed in section 7.2.6. of reference [100]. The method first estimates the neutrino energy using an observable defined such that it is directly coupled to the charged lepton energy. The jet energy is then corrected by adding the previously evaluated neutrino energy. Furthermore, it was also determined in [100], that the heavy quark semileptonic decays affect the precision on the jet energy, i.e., for those cases when they occur, by up to $\sigma_{\mathrm{HeavyQ}} = 15\%$ of the total jet energy.

Each of the four major sources of uncertainty summarised above has an impact on the precision with which the energy of a hadronic jet can be measured. Therefore, the jet energy resolution can be expressed as:

$$\sigma_{\mathrm{E_j}} = \sigma_{\mathrm{Modell.}} \oplus \sigma_{\mathrm{Clust.}} \oplus \sigma_{\mathrm{Det.}} \oplus \sigma_{\mathrm{HeavyQ}} \tag{4.2}$$

The contribution from $\sigma_{\mathrm{Modell.}}$ is typically very small compared to the others and therefore it will not be further taken into account in this thesis.

The uncertainty stemming from the inevitable jet clustering errors, $\sigma_{\text{Clust.}}$, however, is not negligible. Its impact depends significantly on the choice of jet algorithm and the associated parameter values. The level of uncertainty introduced by $\sigma_{\text{Clust.}}$ is presently under study, with the aid of the `TrueJet` [91] approach.

The intrinsic energy resolution of the calorimetric system, i.e., the detector subcomponent where the jet energy is typically measured, has a large contribution to the total jet energy resolution. For instance, the ALEPH experiment at LEP measured the total energy in $Z \to q\overline{q}$ events at a centre-of-mass energy very close to the Z boson mass with a combined ECAL and HCAL precision of $\sigma_{\text{E}_{q\overline{q}}}/\text{E}_{q\overline{q}} = 120\%/\sqrt{E}$ [92]. This combined calorimeter performance is significantly worse than the individual ECAL and HCAL energy resolution values quoted in table 4.1 which were obtained from test-beam data. For this type of events, the poor energy resolution of the ALEPH calorimetric system leads to a relative contribution to the jet energy resolution of about $\sigma_{\text{Det.}} \approx 18\%$ of the jet energy.

A contribution of almost the same order of magnitude, i.e., $\sigma_{\text{HeavyQ.}} \approx$ 10-15% [100] of the jet energy, is caused by the presence of heavy quarks in the hadronic jet and their subsequent semileptonic decays.

4.2 Motivation for a New Approach to Measuring Jet Energies

Many final states relevant for physics studies at the ILC are characterised by a high jet multiplicity plus the presence of charged leptons and/or significant amounts of missing energy. For instance, $t\bar{t}$ events directly produced in e^+e^- collisions, can have up to six jets in the final state or more with gluon radiation.

For the level of precision envisioned at the ILC, the jet energy resolution must at least enable a clear separation between the W and Z boson hadronic decays. This is highly important for ILC physics studies like, e.g., determining the Higgs self-coupling [73], [70].

One of the two processes considered for this purpose is the double Higgs-strahlung: $e^+e^- \to ZHH$. Since the Z boson decays with a branching ratio of $\approx$70% to hadrons, $e^+e^- \to ZHH \to qqbbbb$ becomes one of the most important signal channels. Due to the rather low cross section (figure 7 from [101]) and the complex event topology (6 jets), distinguishing the signal channel from the major backgrounds, $e^+e^- \to t\bar{t} \to bWbW$ and $e^+e^- \to ZZZ$ or $ZZH \to qqbbbb$, becomes a real challenge. This requires a high precision H, W and Z boson separation.

The achievable degree of separation can be evaluated as follows, based on the reconstructed W and Z invariant mass peaks obtained from $W \to qq$ and $Z \to q\bar{q}$ di-jet events:

$$W/Z_{separation} = \frac{M_Z - M_W}{\sigma_{\Delta M}}, \text{ where } \sigma_{\Delta M} = \sqrt{\sigma^2_{M_W} + \sigma^2_{M_Z}} \tag{4.3}$$

and M_Z, M_W represent the respective gauge boson mass, i.e., the mean values of the two di-jet mass distributions. The accuracy with which M_Z and M_W can be

extracted from the distributions is fundamentally limited by the intrinsic natural decay width (Γ) of the gauge bosons.

This establishes the precision goal at the ILC: the mass resolution of each di-jet system, i.e., σ_{M_W}/M_W and σ_{M_Z}/M_Z, respectively, should be at least of the same order as the natural decay width of the gauge bosons, i.e., $\sigma_{M_{jj}}/\mathrm{M}_{jj} \approx \Gamma_W/\mathrm{M}_W \approx \Gamma_Z/\mathrm{M}_Z \approx 0.027$. This translates to: $\sigma_{M_W} = 2.17\,\mathrm{GeV}$ and $\sigma_{M_Z} = 2.46\,\mathrm{GeV}$. Using these values in equation 4.3 shows that the achieved separation between the W and the Z mass peaks would be of the order of $\approx 3.3\sigma$ in this case.

The implications of the above mentioned precision goal for the detector performance and jet energy resolution were analysed [7] in a simplified physics case. For this purpose, only the decays of the Z boson to light quarks were taken into account. Furthermore, it was considered that the Z boson, produced in $e^+e^- \to Z$ collisions at the ILC, decays at rest. Thus, in this particular case, the visible energy measured in the event is equivalent to the di-jet energy and no jet clustering is required. Concerning the jet energy resolution defined by equation 4.2, these simplifications imply that the uncertainties coming from the semileptonic heavy quark decays (σ_{HeavyQ}) as well as from the jet clustering ($\sigma_{\mathrm{Clust.}}$) can be ignored. Consequently, for this type of events, the jet energy resolution can be approximated with:

$$\boxed{\sigma_{\mathrm{E_j}} \approx \sigma_{\mathrm{Det.}}} \tag{4.4}$$

This approximation will be further employed throughout the rest of this thesis.

Since usually the energies of the particles forming the jets, i.e., the jet energies, are measured in the calorimetric system, $\sigma_{\mathrm{Det.}}$ itself can be approximated with $\sigma_{\mathrm{Det.}} \approx \frac{\alpha}{\sqrt{E}} \oplus \beta \oplus \frac{\gamma}{E}$. In the traditional approach to calorimetry, the jet energy resolution is usually dominated by the stochastic term. Therefore, a further simplification is performed by ignoring the constant and noise term: $\sigma_{\mathrm{E_j}} \approx \sigma_{\mathrm{Det.}} \approx \frac{\alpha}{\sqrt{E}}$. Implicitly, the *di-jet* energy resolution is also dominated by α and it affects the di-jet mass resolution with a contribution that can be approximated [7] by:

$$\frac{\sigma_{M_{jj}}}{M_{jj}} \approx \frac{\alpha(E_{\mathrm{jj}})}{\sqrt{E_{jj}}} \tag{4.5}$$

where M_{jj} and E_{jj} represent the mass and the energy of the di-jet system, respectively.

Figure 4.1 shows the typical jet and di-jet energies ($\mathrm{E_{jj}}$) expected in ILC collisions that result in four-jet events, considering a centre-of-mass energy of $\sqrt{s}$=500 GeV.

It can be seen that the typical di-jet energy range at the ILC, running at a centre-of-mass energy of $\sqrt{s}$=500 GeV, expands between $\mathrm{E_{jj}}$=100 GeV and $\mathrm{E_{jj}}$= 300 GeV. Considering the precision requirement on the di-jet *mass* resolution of $\sigma_{M_{jj}}/\mathrm{M}_{jj} \approx 0.027$ and its relation to the di-jet *energy* resolution, expressed in equation 4.5, an upper limit on the stochastic term $\alpha(\mathrm{E_{jj}})$ can be determined. Thus, for low di-jet energies $\alpha(\mathrm{E_{jj}}$=100 GeV) should be smaller than 27% for the combined calorimetric system; while for the higher di-jet energies $\alpha(\mathrm{E_{jj}}$=300 GeV) must not exceed 46%. In the context of the simplifications described above, this

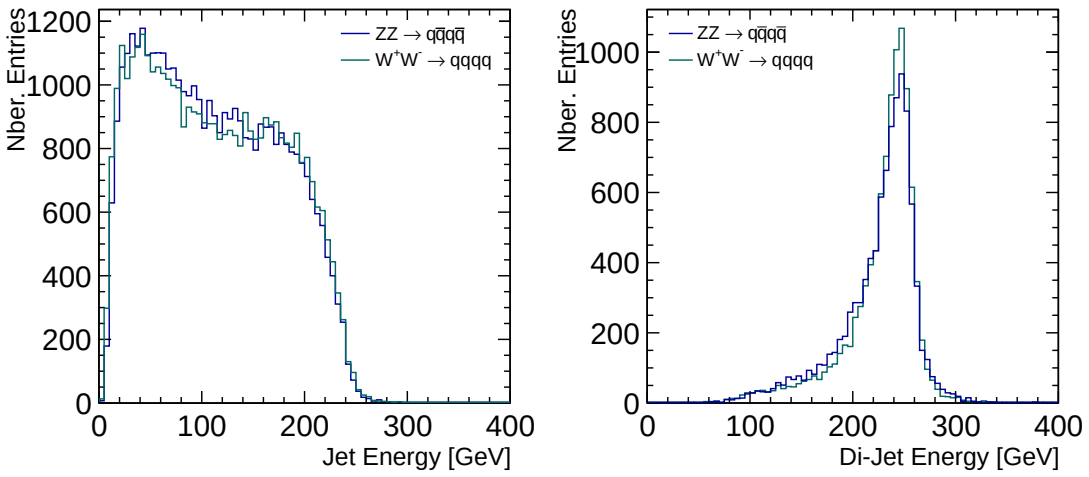


Figure 4.1: Single jet (left) and di-jet (right) energy distributions observed in simulated W^+W^- and ZZ hadronic events at the ILC with a centre-of-mass energy of $\sqrt{s}$=500 GeV. The four jets were clustered using the k_T jet clustering algorithm [102] *after* applying the procedure for removing the $\gamma\gamma \rightarrow$ hadrons background. The jet pairing was performed using information from a kinematic fit.

translates to a requirement that the relative jet energy resolution should not exceed $\sigma_{\mathrm{E_j}}/\mathrm{E_j} < 3\text{-}4\%$ for the range of jet energies that can be reached at the ILC. These values define the precision goals for calorimetry and jet reconstruction at the ILC.

In comparison to the performance achieved by ALEPH, i.e., a value of 120% for the stochastic term and a relative jet energy resolution of approximately 18%, it is evident that the ILC precision goals require a detector energy resolution ($\sigma_{\mathrm{Det.}}$) at least a factor of four better. This level of accuracy cannot be attained following the traditional approach of measuring the particles' energy exclusively in the calorimetric system.

A very promising solution is provided by the *Particle Flow* concept which will be presented in the following.

4.3 The Particle Flow Concept

The Particle Flow approach was developed from a set of fundamental observations that will be discussed in the following. Previous experiments, like, e.g., ALEPH [103], ZEUS [104], H1 [95] or CDF [105], have already used key Particle Flow aspects in their event reconstruction. To illustrate the basic principles, it is useful to consider a typical multijet event like the one presented in figure 4.2. The displayed event shows an $e^+e^- \rightarrow W^+W^- \rightarrow qqqq$ interaction recorded by the ALEPH detector [103] while the LEP accelerator was running at a centre-of-mass energy of $\sqrt{s} \approx$207 GeV.

Due to its four-jets topology, the event is rather busy in terms of tracks and

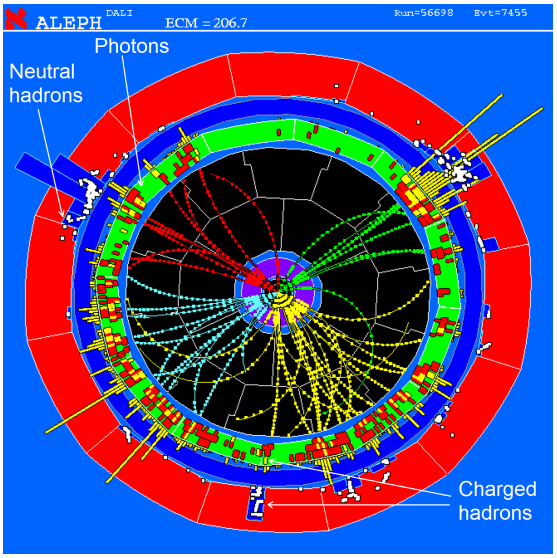


Figure 4.2: Event display showing an $e^+e^- \to W^+W^- \to q\bar{q}q\bar{q}$ event recorded at the ALEPH detector [103] at LEP. Each jet is represented by a different colour. Image taken and modified from [106].

energy deposits in the different sub-detectors. The key to distinguishing between different particle types is the fact that each type elicits a different response from each sub-detector it passes.

The first LEP investigations of the W and Z hadronic decays provided crucial information regarding the typical energy content of a jet. Thus, about 25-27% of the jet energy is deposited by photons (most of them from π^0 decays), 62-65% are charged particles (hadrons, electrons and muons), the neutral hadrons ($\mathrm{K}_L^0, n, \bar{n}$) form approximately 10% of the jet energy while about 1.5% is carried away by neutrinos, when no heavy quarks are present [107]. These percentages must be interpreted as average expectation values. As mentioned earlier, the individual jet content and multiplicity can vary significantly from jet to jet.

The photons and electrons typically deposit almost their entire energy in the ECAL (bright green in figure 4.2). The charged and neutral hadrons interact both in the ECAL and in the HCAL (red). At the same time, the charged particles obviously leave tracks in the volume of the tracking system (black) as well.

Generally, the measurement of a charged particle's transverse momentum (P_T), carried out in the tracking system, is significantly more precise than the direct energy measurement performed by the calorimeters. For example, the resolution achieved by the ALEPH tracker was $\sigma(1/P_T) = 1.6 \cdot 10^{-4}\,\mathrm{GeV}^{-1}$ [92]. Furthermore, the typical HCAL resolution is notably worse than the ECAL precision, as indicated by the values in table 4.1. More importantly, the energy of a charged particle can be determined based on the tracker information.

The traditional approach to reconstructing jets relies entirely on the ECAL and/or HCAL measurements. Consequently, in the case of a multi-jet event, like the one shown in figure 4.2, up to ≈75% of the jet energy is measured in the HCAL. This estimation is based on the particle composition of the typical jet mentioned above [107]. Thus, the largest fraction of a jet's energy is measured in

the sub-detector with a rather poorer energy resolution. This situation is illustrated in the left pannel of figure 4.3.

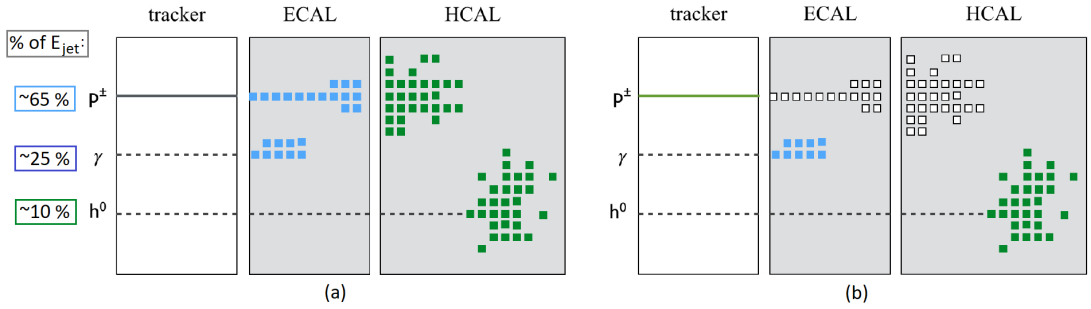


Figure 4.3: Sketch illustrating the Particle Flow concept. Pannel (a) shows the traditional approach relying entirely on the calorimetric measurements for the energy determination. In contrast, pannel (b) presents the Particle Flow idea. Figure taken from [108].

In contrast, the Particle Flow concept relies on the idea that *every individual particle* produced in the initial interaction should be measured in the sub-detector that offers the best (energy) resolution for that particle type. This approach is presented in pannel (b) of figure 4.3:

- → The *charged particles*, constituting about 62-65% of the jet energy, should be measured exclusively in the tracker. The information regarding the clusters they created in the calorimetric system is discarded.
- → The *photons* which account for approximately 25-27% of the jet energy should be measured in the electromagnetic calorimeter.
- → Only ≈10% of the jet energy carried by the *neutral hadrons* should be measured in the hadronic calorimeter.

Hence, in comparison to the traditional approach, the Particle Flow paradigm foresees that only a very small fraction of the jet energy is measured in the sub-detector with the worse energy resolution. This is expected to significantly improve the achievable jet energy resolution.

Assuming, for example, some presently achievable ECAL ($15\%/\sqrt{E(GeV)}$) and HCAL ($55\%/\sqrt{E(GeV)}$) energy resolutions, a Particle Flow jet energy resolution of $\sigma_{\mathrm{E_{jet}}}/\mathrm{E_{jet}}=19\%/\sqrt{E}$ could be attained [7]. However, this estimation considers that the energy deposits created by a particle throughout the detector are always perfectly asociated to the right particle. In reality, this is clearly not the case. Especially in a busy event like the one illustrated in figure 4.2, it is not always posible to disentangle energy deposits in the calorimeters and correctly associate the created clusters to the particles that produced them. Occasionally, clusters can be inadvertently split or fragments from different clusters may wrongly be grouped together. Moreover, wrong associations between tracks and clusters can also occur. This is particularly relevant for the Particle Flow approach since the cluster information is rejected when determining the energy of charged particles.

Therefore, these confusion errors constitute a challenge for the Particle Flow performance.

In view of these considerations, the jet energy resolution approximation expressed in equation 4.4 must be readjusted to reflect the contribution from the unavoidable Particle Flow confusion:

$$\sigma_{\mathrm{E_j}} \approx \sigma_{\mathrm{Det.}} \oplus \sigma_{\mathrm{Conf.}} \tag{4.6}$$

Studying the impact of $\sigma_{\mathrm{Conf.}}$ in the context of a dedicated physics analysis is one of the main tasks of this thesis.

Despite the introduction of the $\sigma_{\mathrm{Conf.}}$ term, the major improvement provided by the Particle Flow philosophy consists of significantly reducing the contribution of $\sigma_{\mathrm{Det.}}$ to the jet energy resolution. In contrast to the traditional approach, under the Particle Flow paradigm, the detector precision is no longer dominated by the large uncertainty of the calorimeteric measurements. Thus, $\sigma_{\mathrm{Det.}}$ must be reformulated as:

$$\sigma_{\mathrm{Det.}} = f_{h^{\pm}} \cdot \sigma_{tracker} \oplus f_{\gamma} \cdot \sigma_{ECAL} \oplus f_{h^0} \cdot \sigma_{HCAL} \tag{4.7}$$

where f_i represents the fraction of the total jet energy carried by charged particles ($i = h^{\pm}$), photons ($i = \gamma$) and neutral hadrons ($i = h^0$), while σ_{ECAL} and σ_{HCAL} denote the electromagnetic and hadronic single particle energy resolutions., defined by relation 4.1.

In order to fulfill the ILC precision goal for the jet energy resolution, all factors in equation 4.7 must be as small as possible. This requirement has a major impact on the design of a Particle Flow detector as will be discussed in the following.

4.4 Implications for Detector Design

The Particle Flow concept aims at improving the jet energy resolution by measuring each outgoing particle using the most adequate detector component.

Regardless of the approach, it is highly desirable that each sub-detector provides the best achievable intrinsic resolution. In addition, for Particle Flow it is essential that the contribution from the mis-associations term (σ_{Conf} in equation 4.6), also known as the *confusion term*, to the jet energy resolution is minimal as well. To achieve this goal, it is crucial to distinguish accurately the interactions in the detector material created by different particles. This latter consideration enforces the following requirements on the detector design. The subsequent discussion follows the concepts, studies and results presented in [7] and [109].

- **High three-dimensional ECAL and HCAL granularity**: It is very important for the Particle Flow performance to distinguish as well as possible: *(i)* between energy deposits belonging to two (or more) different calorimetric showers and *(ii)* between electromagnetic and hadronic showers. The first condition in particular places a strong requirement for the detector

granularity. Thus, the calorimeter design should entail a material characterised by a small Molière radius[2] and a high longitudinal and transverse segmentation. Ideally, the size of a calorimeter cell should be much smaller than the Molière radius. The high granularity of the calorimeters allows a more accurate measurement of the individual energy deposits and hence a significantly better separation of the calorimetric showers of different origin.

- **Large detector size:** The separation of the calorimetric energy deposits themselves is better the larger the size of the detector, i.e., it scales with the inner radius of the ECAL in the barrel and with the detector length along the z-axis in the endcap region.
- **Large axial magnetic field:** A high magnetic field is desirable since it deflects the charged particles away from a jet leaving only the neutral particles to travel on a straight line trajectory. This can improve the particle shower separation and association. Furthermore, both the ECAL and HCAL should be inside the solenoid.
- **Little material preceding calorimeters:** For a good Particle Flow performance, the particles produced in the collision should suffer the minimum possible amount of energy loss in the other sub-detectors they traverse before reaching the calorimeters.
- **Good hermeticity:** Since the aim of Particle Flow is to measure every visible particle in an event, it is important to ensure that most if not all particles stand a chance of being detected. This implies that the detector design should forsee as few gaps and non-instrumented areas as possible.

These prerequisites are very important for minimising the potential association errors. Thus, they ensure that the cluster and cluster-track mis-associations do not severely degrade the jet energy resolution. Consequently, a detector designed according to these principles is highly compatible with the Particle Flow philosophy.

Several past and on-going experiments have used different variations of Particle Flow to improve their energy measurements. Their approach and performance offer valuable lessons in terms of detector and algorithm design. These aspects will be presented in the following.

4.5 Overview of Past and Present Implementations

The first steps towards a Particle Flow approach were made by the experiments at the Large Electron-Positron Collider (LEP), more than 25 years ago. Known

[2]The Molière radius is a material constant that describes the transversal size of a fully contained calorimetric shower. It is defined as the radius of a virtual cylinder that contains, on average, 90% of the energy deposited by the shower.

as "Energy Flow" at the time, the method delivered promising results, as will be discussed in the following. Since then, subsequent experiments, with different detector designs, have adopted it as part of their standard reconstruction.

For the precision requirements at the planned ILC, the Particle Flow approach is paramount. The present section briefly describes the historical context that informed the manner in which Particle Flow will be implemented at the ILC. For this purpose the following experiments, each an example from a presently running or recent high-energy collider, were considered:

- ⇨ **ALEPH**: [103] was a multi-purpose particle detector which measured collision events at LEP at a centre-of-mass energy of $\sqrt{s}$ ≈91 GeV between 1989 and 1995 and then at $\sqrt{s}$ ≈200 GeV between 1995 until 2000.

- ⇨ **ZEUS**: [104] ran from 1992 until 2007 at the Hadron-Electron Ring Accelerator (HERA), located at DESY Hamburg. Its main goal was to study the internal structure of the proton and perform tests of the Standard Model with the aim of searching for new physics. The ZEUS detector measured deep inelastic scattering events produced at HERA by colliding electrons (or positrons) with protons at a centre-of-mass energy of $\sqrt{s}$=318 GeV.

- ⇨ **CDF**: [105] the Collider Detector at Fermilab (CDF) was an experiment located at the Fermi National Accelerator Laboratory in the United States. It was operated from 1985 until 2011 and it recorded proton-antiproton collision events produced at a centre-of-mass energy of up to $\sqrt{s}$=1.8 TeV by the Tevatron Collider.

- ⇨ **CMS**: [48] The Compact Muon Solenoid (CMS) is one of the two general purpose experiments currently ongoing at the Large Hadron Collider (LHC) accelerator at CERN. The LHC collides two beams of protons reaching, at the time of writing, an unprecedented centre-of-mass energy of $\sqrt{s}$=13 TeV.

All four detectors follow a similar design, illustrated by the layout depictions in figure 4.4. It follows the typical "onion-like" multilayered approach with a set of vertex and tracking detectors in the immediate vicinity of the interaction point (IP) followed by a large central tracker, an electromagnetic and a hadronic calorimeter, the latter positioned either within or outside of the magnetic coil. The last layers typically comprise the muon chambers.

To appreciate the specifics of how each experiment has applied the Particle Flow concept, it is important to ascertain the compatibility of their respective detector design with the requirements presented in the previous section.

4.5.1 Compatibility with the Particle Flow Detector Requirements

While a detailed description of the four Particle Flow pioneering experiments is beyond the scope of this section, their most relevant detector design aspects are summarised in table 4.2 and discussed in the following.

Design aspect:		ALEPH	ZEUS	CDF	CMS
Magnetic field [T]		1.5	1.43	1.4	4
Tracker type		TPC	wire drift chamber	wire drift chamber	Si strips
Amount of material before calorim. [%X_0]		7.1	≈100	7	30
ECAL	Position w.r.t. solenoid	inside	outside	outside	inside
	Absorber	Pb	Depleted U	Pb	$PbWO_4$
	Detector	Wire chamber	Scintill. plates	Scintill. plates	$PbWO_4$
	Total X_0 [a]	22	21	19	25.8
	Inner R [m]	1.85	1.23	1.73	1.3
	Transv. granul.	3×3 cm^2	5×20 cm^2	5×25 cm^2	2×2 cm^2
	Longit. granul.	3 layers	-	-	-
HCAL	Position w.r.t. solenoid	outside	outside	outside	inside and outside
	Absorber	Fe	Depleted U	Fe	Brass
	Detector	Streamer tubes	Scintill. plates	Scintill. plates	Scintill. plates
	Total λ_I [b]	7.2	4.1	4.5	5.5-10
	Transv. granul.	12×12 cm^2	10×20 cm^2	$\Delta\phi \times \Delta\eta$=15°×0.1	10×10 cm^2
	Longit. granul.	-	2 layers	-	-

Table 4.2: Comparison of the design characteristics relevant for the Particle Flow implementation. The quoted values were taken from: [103] (ALEPH), [104], [93], [94] (ZEUS), [105], [114], [115] (CDF) and [48] (CMS).

[a] The radiation length, X_0, is a material constant that expresses the mean distance over which the energy of an electron decreases by a factor of 1/e as a result of radiation loss [113]

[b] The nuclear interaction length, λ_I, is defined as the average distance that a hadron can travel through the material without suffering an inelastic nuclear interaction.

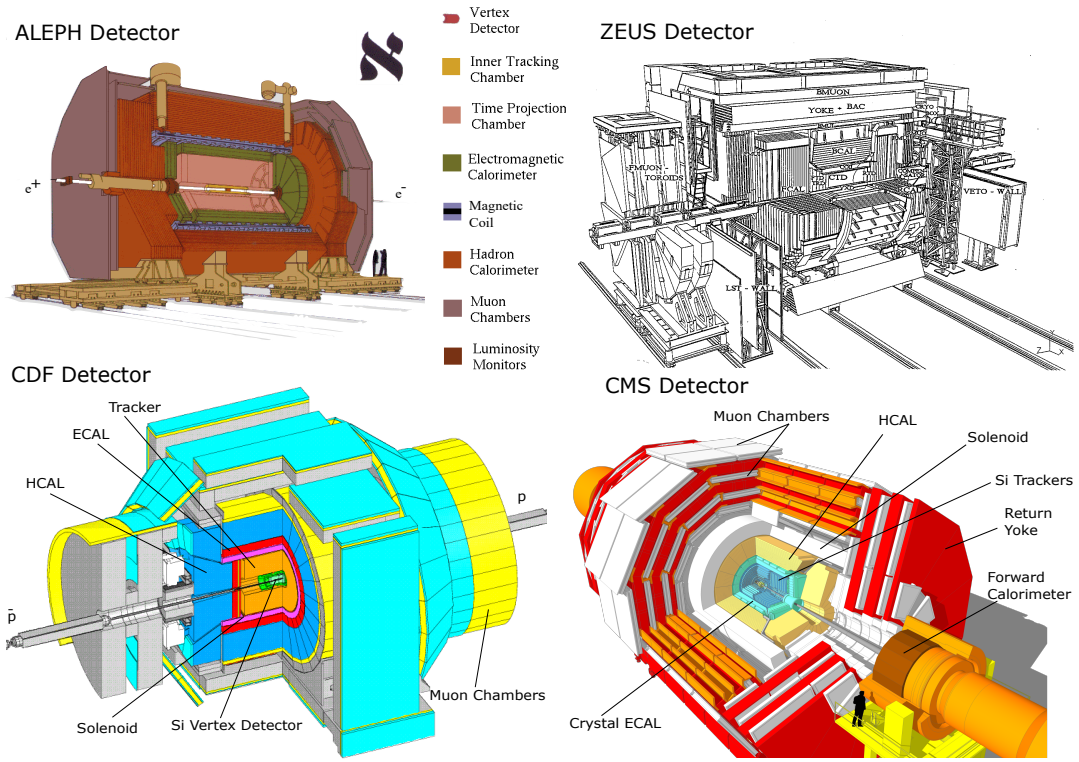


Figure 4.4: Sketches illustrating the layout of four detectors that have used a Particle Flow approach in their event reconstruction. From left to right: the ALEPH [110] and ZEUS [104] detectors on the first row, the CDF [111] and CMS [112] detectors on the lower row. Figures taken from the mentioned references.

High granularity: In terms of transverse granularity, from the four considered detectors, the CMS electromagnetic and hadronic calorimeters have the smallest cell sizes. Furthermore, the CMS ECAL also fulfills the Particle Flow requirement that the cell size should be of the same order as the Molière radius, with $R_M(PbWO_4) \approx 2\,\text{cm}$. However, neither detector presents a high longitudinal granularity, with only the ALEPH ECAL and the ZEUS HCAL providing more than one measurements along the calorimeter tower.

Calorimeter types: Concerning the class of calorimeter, all four experiments employed sampling electromagnetic and hadronic calorimeters with the exception of CMS which uses a homogeneous ECAL. While sampling calorimeters are more cost effective and can benefit from a longitudinal granularity, using a single block of material to serve as both absorber and active medium also has advantages. Firstly, no shower particles would be missed by interacting in the absorber before detection and, secondly, the detector response is the same everywhere, leading to a superior energy resolution.

Due to their absorber material, consisting almost entirely of uranium (98.1% U^{238}, 1.7% Nb, 0.2% U^{235}) [104], the ZEUS calorimeters were energy *compensating* and, hence, especially interesting.

Electrons and photons deposit their energy in the calorimeter material differ-

ently from hadrons. However, in measuring jet energies, it is very important that the calorimeter response is as independent as possible of the hadronic or electromagnetic nature of the jet particles. The ratio between the electromagnetic and hadronic signal response is typically e/h≈1.1-1.35 in a non-compensating detector. In contrast, a compensating calorimeter would provide an electron/hadron signal ratio close to one, i.e., e/h≈1.

In uranium, slow moving neutrons were produced as π^0 particles interacted with the calorimeter absorber. The neutrons would then be detected by the scintillator tiles, thus increasing the hadronic signal.

Consequently, the single particle energy resolution of the ZEUS calorimeter was determined to be $\sigma(E)/E=18\%/\sqrt{E}$ for electromagnetic particles and $\sigma(E)/E=35\%/\sqrt{E}$ for hadrons [104]. In contrast to the other energy resolution values quoted in table 4.1, the performance achieved by the ZEUS detector is the closest to the ILC goals. However, utilising depleted uranium as absorber material is nowadays no longer feasible. Furthermore, despite its energy compensating nature, the ganularity of the ZEUS calorimeters is rather low and hence insufficient to fulfil the ILC precision requirements.

Magnetic field and detector size: Usually, a large detector radius is beneficial for Particle Flow as it improves the separation of energy deposits in the calorimeters [7]. The size of the ECAL inner radius is comparable in all four detectors, with ALEPH being the largest, at almost two metres.

The strength of the axial magnetic field is also important for the same purpose. It is very similar for ZEUS, CDF and ALEPH, while in the case of CMS the solenoid gives rise to a B field almost three times stronger.

Material preceding calorimeters: In order to minimise the effects (e.g., bremsstrahlung, etc.) that can lead to energy loss, the amount of material the outgoing particles must cross before reaching the calorimeters should be as small as possible. The ALEPH and CDF detectors are closest to fulfilling this requirement. In comparison, since the ZEUS calorimeters are located outside of the solenoid, the particles must pass through almost an entire radiation length of material before reaching them.

Hermeticity: It is important for the Particle Flow philosophy that as few particles as possible escape detection through cracks in the detector material or uninstrumented areas. All four detector designs are compliant with this requirement, offering coverage of almost 4π sr. For example, in the case of ALEPH, the gaps correspond to only 2% of the total material in the barrel and 6% in the end-caps.

From the construction aspects discussed above and summarised in table 4.2, it is clear that, while each detector fulfills a fraction of the Particle Flow requirements, none of them was specifically designed for it. Consequently, it is very useful to examine the manner in which each experiment has exploited their relevant design features to benefit from implementing the Particle Flow concept in

their reconstruction. These matters will be discussed in the following.

4.5.2 Implementation of the Particle Flow Algorithm

The Particle Flow paradigm influences not only the key aspects of detector design but also the manner in which collision events are reconstructed.

For instance, the energy measurement of a charged particle is meant to be carried out based only on the tracker information. The calorimeter measurements associated to the charged particle are no longer used for the energy determination. In this context, it is crucial that the right calorimetric information is discarded. Consequently, the interplay and correct association between the measurements performed by the different sub-detectors is very important. In this regard, the Particle Flow reconstruction is essentially a very complex pattern recognition algorithm.

The manner in which the previously mentioned four experiments, ALEPH, CDF, ZEUS and CMS, have implemented the Particle Flow concepts in their reconstruction algorithms is briefly presented in the following.

ALEPH: The ALEPH approach to realising Particle Flow [92] set the standard for its implementation. The tracks that have passed a pre-selection stage are extrapolated to the calorimeter interface. Taking into account the tracks' orientation and the position of the nearby clusters, the tracks and clusters are associated into groups of so called "*calorimeter objects*".

Each one of the latter is processed according to the following sequential identification procedure. Firstly, the energy of all surviving tracks is measured and counted as *charged energy* by making the mass assumption that all the charged particles creating tracks are pions. Secondly, a set of identification procedures is carried out to identifiy muons, electrons and photons. The calorimetric energy associated to the photons is taken into account as *electromagnetic neutral energy*. After each identification step, the corresponding energy deposits are removed from the calorimeter objects. The remaining calorimetric energy that is not accounted for is added up. The sum is compared to the energy of any tracks remaining in the calorimeter object. If it either exceeds the total track energy by 500 MeV or no track is left in the calorimeter object, then the calorimetric energy is classified as *neutral hadronic energy*. These processing steps are repeated for all calorimeter objects. The result comes in the form of a list of so called "*energy flow objects*", i.e., a list of particles - electrons, muons, photons, charged and neutral hadrons.

ZEUS: The ZEUS detector had the advantage of a compensating calorimeter which provided a very good energy resolution for hadrons. Therefore, in implementing a Particle Flow algorithm [116], the tracker measurement was used instead of the calorimeter one *only* in those cases when the resolution of the former was even better. In order to decide when the tracker information should be used, the level of precision obtained with the two sub-detectors was compared on a case by case basis. For this purpose, it was essential to ensure that the calorimeter

energy deposits associated to a track come only from the respective track alone. If the resolution obtained from the track measurement was better than the energy resolution of the calorimeter, the latter was discarded.

CDF: At CDF, due to the rather coarse transverse segmentation, the event reconstruction was carried out based on detector radial segments called "towers". The principle guiding the Particle Flow implementation [117] was to make better use of the full detector granularity by performing energy corrections on a tower level. Consequently, the track momentum was used in order to reduce unwanted effects (e.g., the calorimeter nonlinearity) in the calorimetric measurement. Furthermore, the shower position information measured at the average shower maximum in the ECAL, was also employed in order to separate overlapping electromagnetic showers, e.g., like in the case of π^0 and $\pi^\pm$. Finally, the towers determined in the reconstruction were divided into four categories: *(i)* track towers containing a track but no associated cluster, *(ii)* gamma towers consisting of a cluster but no associated track, *(iii)* mixed towers comprising both a cluster and track and *(iv)* towers that could not be assigned containing energy deposits that could form neither a track nor cluster. The latter took up only 3.4% of the event energy while the mixed category accounted for up to 58% [118]. For each tower category, a different method was then used in order to determine the energy of the tower, according to the type of particles giving rise to it.

CMS: The CMS Particle Flow algorithm [119] is similar to the ALEPH implementation. Taking advantage of the higher transverse granularity of the detector, the CMS approach first reconstructs tracks and clusters with a high efficiency even in a complex collision environment. A *linking algorithm* is then run for each pair of elements, i.e., cluster-track, cluster-cluster and track-track in the event. The output is a number of "blocks" which contain clusters and/or tracks that are linked. The usual number of clusters and/or tracks contained in a block ranges from one to three. An identification sequence analogous to ALEPH is then carried out to identify electrons and muons. In the following step, the momentum of the track(s) and the calorimetric measurements contained in a block are compared. If the total energy of the ECAL and HCAL clusters in a block is significantly larger than the total momentum of the associated tracks, a "*particle flow photon*" and/or a "*particle flow neutral hadron*", respectively, are created from the surplus energy. When the energy measured in the calorimeters is at least three standard deviations smaller than the associated track momentum, a new procedure that aims at identifying muons and fake tracks is applied. The successfully associated track-cluster pairs are classified as charged hadrons and their energy is determined from the track measurements.

Despite the fact that none of the detectors considered in this historical overview was specifically designed nor optimised for Particle Flow, the implementation of the new paradigm in their reconstruction procedure provided promising results. They are discussed in the following section.

4.5.3 Performance

The results and improvements achieved by the ALEPH, ZEUS, CDF and CMS experiments by utilising the Particle Flow approach are summarised below.

ALEPH: The mass resolution as a function of the invariant mass was studied at ALEPH [92] using a sample of observed hadronic ($q\bar{q}$) events that have an isolated hard photon (or neutral pion) produced either as initial or final state radiation. The signal selection criteria required the photon to have an energy larger than 20 GeV and to be well isolated. Furthemore, the invariant mass of the photon-jet system was required to be larger than 10 GeV.

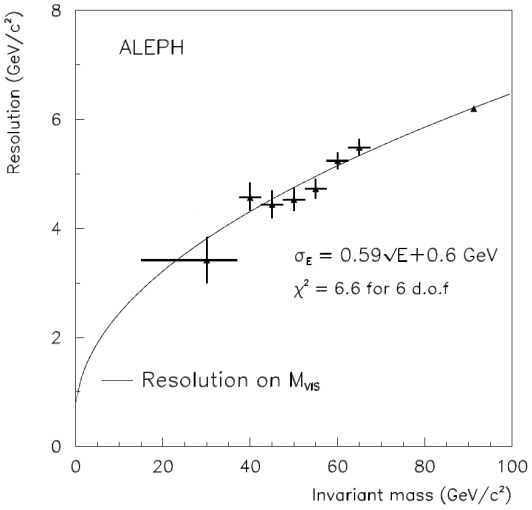


Figure 4.5: The di-jet mass resolution in $q\bar{q}$ events observed at ALEPH [103] expressed as a function of the invariant mass. Figure adapted from [92]

The result of this study is presented in figure 4.5 as the resolution of the invariant di-jet mass versus the invariant mass, where $M_{q\bar{q}}$ is measured directly with the Particle Flow procedure (M_{VIS}) and is denoted by the data points.

Considering equation 4.5, the resolution curve illustrated in figure 4.5 can implicitly be interpreted as a function of $\sqrt{E_{q\bar{q}}}$. The continuous line in figure 4.5 represents the fit to the M_{VIS} data points of a typical jet energy resolution, dominated by the calorimeter performance. The fit result leads to the final expression for the resolution: $\sigma_{M_{q\bar{q}}}/M_{q\bar{q}} \approx \alpha(E_{q\bar{q}})/\sqrt{E_{q\bar{q}}} = (59\pm3)\%\sqrt{E} + (0.6\pm0.3)$ [92].

In contrast, the energy resolution achieved at ALEPH, using the traditional approach for $Z \rightarrow q\bar{q}$ events at a centre-of-mass energy close to the Z mass, was $\sigma(\mathrm{E})/\mathrm{E} = 120\%/\sqrt{E}$ [92]. Therefore, it can be surmised that the implementation of Particle Flow at ALEPH led to approximately a factor two improvement in the jet energy resolution.

However, it can be seen from figure 4.5 that at invariant mass values close to the W and Z masses, i.e., 80-90 GeV, the achieved di-jet mass resolution is of the order of $\approx$ 6 GeV. This is almost a factor two to three worse than the ILC requirement for the di-jet mass resolution to be of the same magnitude as

the natural decay width (Γ) of the gauge bosons. The need for improvement is also illustrated by the comparison between the magnitude of the stochastic term obtained by ALEPH with Particle Flow and the ILC requirement of 30%.

ZEUS: The performance of the ZEUS Particle Flow algorithm was studied [116] using simulated diffractive events depicted in figure 4.6a. For this purpose, the relevant observable was the ratio ($E_{measured}/E_{gen}$) between the total energy measured in the event, after the reconstruction stage, and the initial energy on generator level. Figure 4.6 shows two ratio distributions plotted versus the generated energy.

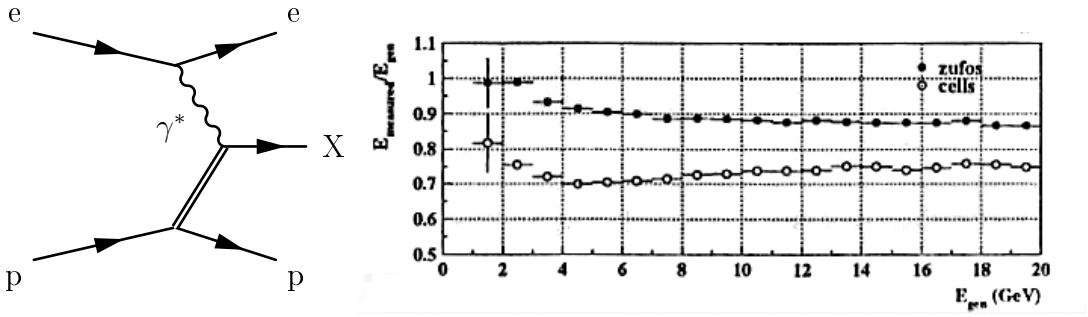


Figure 4.6: Improvements obtained by employing the Particle Flow technique at ZEUS (b) on simulated diffractive events depicted in (a). Plot taken from [116].

In the case of the first distribution, marked by solid circles ("zufos"), the numerator of the ratio ($E_{measured}/E_{gen}$) was determined from the Particle Flow output. In comparison, for the lower distribution, depicted by open circles ("cells"), the numerator was obtained using the traditional calorimetric approach. The comparison was performed using the same events, therefore the denominator was the same in both cases.

It can be seen from figure 4.6 that the Particle Flow reconstruction is much closer to the ideal ratio value of 1. On average, using the traditional method, only 72% of the generated energy could be reconstructed as opposed to the Particle Flow procedure that could reconstruct 90% of the initial energy. Furthermore, using the Energy Flow approach enabled the recovery of approximately 70% of the energy lost in the inactive material present in front of the ZEUS calorimeter.

CDF: The CDF implementation of the Particle Flow procedure was compared [118] to their standard method for reconstructing the jet energy, described in [120]. The comparison was performed on real *direct photon production* events recorded during the first run of data taking. At the typical Tevatron energy range, the dominant direct photon production process is the gluon-quark Compton scattering, illustrated in figure 4.7a. In this type of events, the photon is usually well reconstructed in the CDF detector and the jet transverse momentum balances the photon p_T.

The results of the comparison study are shown in figure 4.7b in terms of the jet energy resolution displayed as a function of the photon transverse momentum. It

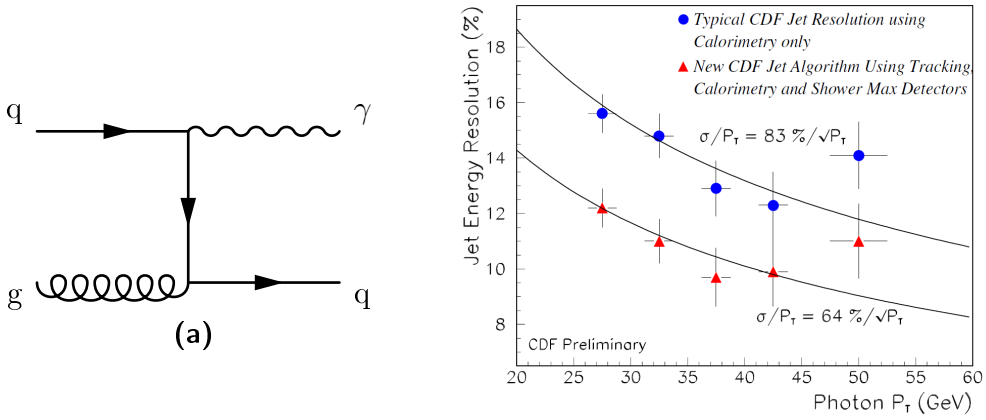


Figure 4.7: Performance of the Energy Flow algorithm (b) applied on direct photon production events sketched in (a). Figure (b) was taken from [118].

can be seen that the jet energy resolution obtained with the new Particle Flow based method (marked with the red triangles) is approximately 20% better than the standard CDF procedure (shown in blue circles). Nevertheless, in comparison to the ILC requirements, the jet energy resolution achieved by CDF is significantly less accurate.

CMS: The performance of the Particle Flow algorithm employed by the CMS collaboration was estimated both on Monte Carlo [119] and on real collision data [121], [122]. The simulation data study was performed with hadronic multijet events and concluded that the jet energy resolution was improved by a factor of 2 to 3 when employing the Particle Flow algorithm [122].

For the studies based on real measured data, p-p collision events observed by CMS in the 2010 run at $\sqrt{s}$=7 TeV were considered. The total number of events corresponds to an integrated luminosity of 35.6 pb^{-1}. The jet energy resolution (in fact, jet p_T resolution) was determined only from di-jet events. A detailed account of the analysis is given in [121].

Figure 4.8a shows the p_T resolution when using CaloJets, while 4.8b presents the result when using Particle Flow jets. The plots also include the jet p_T resolutions (in red) determined from the Monte Carlo truth information from simulated di-jet events. The so called "*Calo Jets*" consist of calorimeter towers formed from combining the energy deposits reconstructed only in the CMS ECAL and HCAL. In contrast, the "*PF Jets*" are formed by running the anti-k_T jet algorithm, with a jet radius size parameter equal to R=0.5, on the list of reconstructed particles provided as output by the Particle Flow procedure.

Comparing the results from the two panels, it is visbile that the Particle Flow reconstruction (4.8b) provides a significant improvement in the lower p_T regions, while at higher p_T the CaloJets and PF Jets resolutions are comparable.

The Particle Flow approach has provided significant improvements, e.g., by a factor two at ALEPH, to the jet energy resolution attained by each of the four

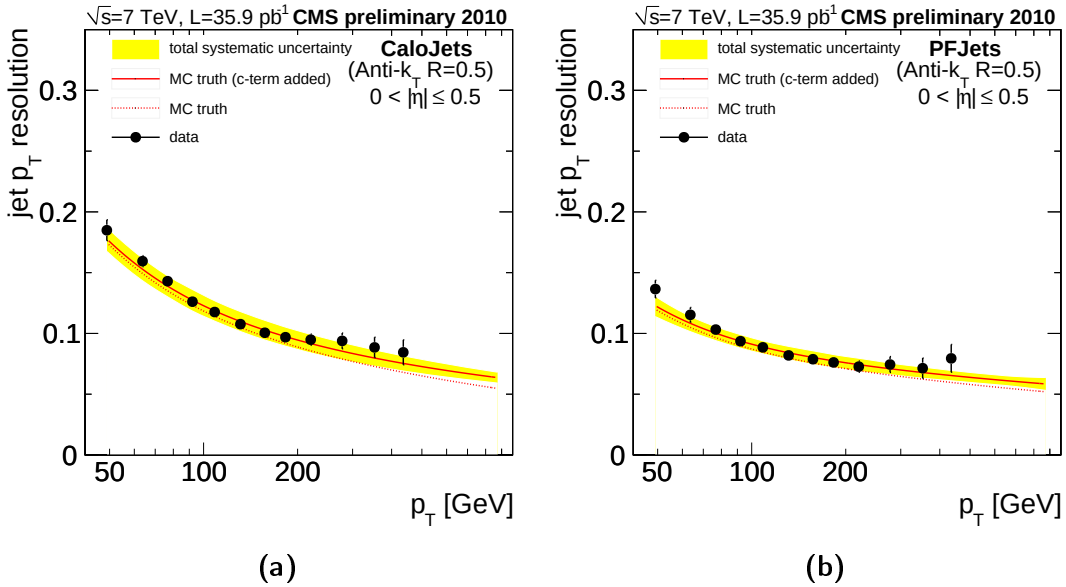


Figure 4.8: The jet p_T resolutions determined for jets obtained only from calorimetric measurements shown in (a) and for jets obtaind from clustering Particle Flow objects in (b). The shown results were obtained within a pseudorapidity (η) range corresponding to the barrel part of the detector. Plots taken from [121].

experiments considered. The outcome was very promising even though none of the detectors had been specifically designed to fullfil the Particle Flow requirements. Especially the granularity prerequisites, crucial for reducing the confusion term in the jet energy resolution parametrisation, were not met.

Therefore, in order to fully benefit from implementing the new paradigm and thus achieve the desired precision goals, the detector design at the ILC must be entirely compatible with Particle Flow. This endeavour is presented in the next chapter in the context of the International Large Detector (ILD) concept.

CHAPTER 5

The International Large Detector Concept

In the quest for a detector design that fulfils all the Particle Flow requirements, two different concepts have emerged for ILC: the *International Large Detector* (ILD) and the *Silicon Detector* (SiD). In contrast to the four detectors presented as study case in section 4.5, the ILD and SiD are specifically optimised for Particle Flow. It is forseen that both detectors will operate in turn at the ILC, in a so called push-pull system.

The SiD concept relies on a compact detector that incorporates a silicon-tungsten electromagnetic calorimeter and a full silicon tracking system. A detailed description of the SiD design concept can be found in [123].

The physics studies presented in this thesis were carried out in the context of the International Large Detector. Therefore, the most relevant aspects of the ILD design concept form the focus of this chapter.

To fully benefit from the Particle Flow approach, all the detector components must provide excellent intrinsic resolutions. Consequently, the ILD design has been extensively optimised through a large number of test-beam and `GEANT4`-based full simulation studies. The results and technical solutions have been first presented in the ILD *Letter of Intent* document [109].

Detailed aspects of the initial design have been further improved and updated in the fourth volume ("Detectors") of the ILC *Technical Design Report* (TDR) [123]. The ILD overall performance and that of its sub-detectors have been exhaustively evaluated for each of the two documents, using two subsequent implementations of the full simulation. The outcome of the evaluations will be discussed in chapter 6. The present overview follows exclusively the detailed description and technical information presented in the TDR.

The ILD is a multi-purpose detector with a length of approximately 13 m and a radius of 7.8 m, operating in an axial magnetic field of 3.5 T. Figure 5.1 presents a transverse section of the ILD, highlighting its global dimensions and the layout of the sub-detectors. They will be briefly presented in the following.

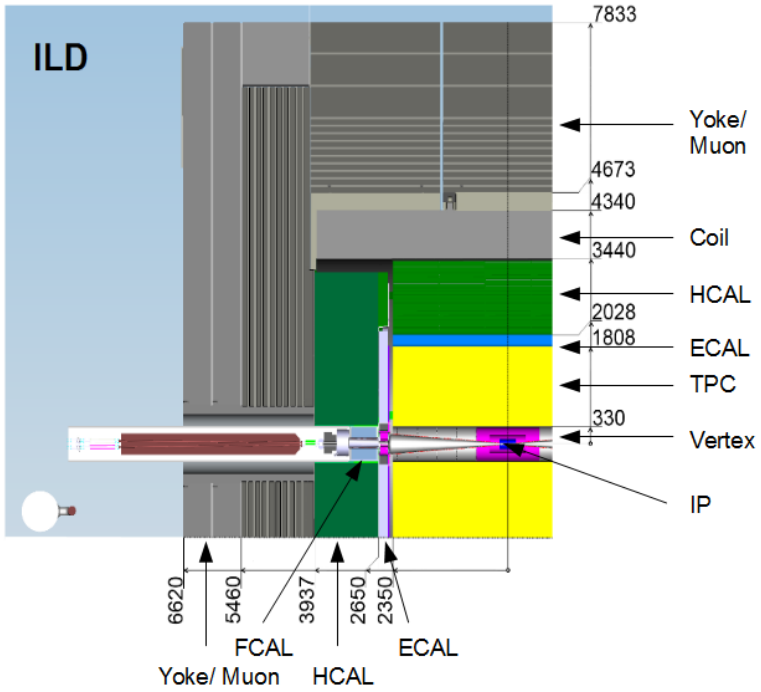


Figure 5.1: Design sketch illustrating a quadrant of the International Large Detector, showing the detector sub-components and their dimensions. Figure taken from [123].

5.1 The Tracking System

The design of the ILD tracking system relies on the complementarity of its components. The central feature is a large time projection chamber (TPC) which constitutes the main tracker. In addition, a silicon-based vertex detector (VTX) provides information regarding tracks with very low transverse momentum and contributes to the identification of secondary vertices. The forward tracker (FTD) extends the coverage of very forward tracks even further, to an angle of approximately 7° with respect to the beam axis. Lastly, a silicon envelope-like system (SIT and SET) surrounds the interaction point and the TPC, respectively, providing not only track measurments but also time-stamping information.

5.1.1 Vertex Detector

The vertex detector plays a major role in detecting low p_T tracks and achieving a high performance identification of heavy flavour quarks and tau leptons. Regarding the latter, the main goal is the precise reconstruction of displaced vertices produced by short lived particles. To reach the level of precision envisioned at the ILC, an impact parameter[1] resolution of the order of: $\sigma_{R-\phi} = 5 \oplus 10/[p \sin^{3/2}(\theta)]$ μm is required.

The design and optimisation of the vertex detector were dictated by the performance considerations summarised in the following:

[1]The impact parameter is the transverse distance, in the $R-\phi$ plane, between the point of closest approach of the particle's trajectory and the interaction point.

- **Single point spatial resolution:** For an accurate track measurement in the vicinity of the interaction point (IP), the single point resolution must be better than 3 μm. This requirement contributes to reducing the association errors that can occur in the Particle Flow reconstruction and, therefore, is very important for separating neighbouring tracks in a jet.
- **Position of first measurement:** In order to achieve the desired flavour tagging performance, the first measurement of a track should be made as close as possible to the IP, i.e., at a distance of a few millimetres.
- **Material budget:** A high number of interactions with the detector's material can lead to the production of secondary particles and, subsequently, affect the Particle Flow reconstruction by increasing the level of confusion. To minimise this, the material budget must be below 0.15% of radiation length per layer.
- **Occupancy:** In order to reduce the potential confusion in track separation and vertex finding, the occupancy of the detector's pixels must be lower than a few percent.

The current design of the ILD vertex detector is based on a cylindrical structure consisting of three, double-sided and concentric sensitive layers. Each layer contains pixel sensors on both sides.

The distance between two sensors, located on opposite sides of the same layer, is approximately 2 mm. Thus, for a particle passing through all three layers, six measurements can be performed. The material budget allocated to each of the three layers is of the order of 0.15% per layer.

The radial distance (measured from the IP) occupied by the vertex detector stretches between 16 mm to 60 mm, i.e., the whole detector thickness is $\approx$5 cm. Furthermore, the first track measurment is performed at a (straight line) distance of 16 mm from the IP.

Three possible options are currently under consideration for the sensor technology: *(i)* CMOS Pixel Sensors [124], *(ii)* Fine Pixel CCD Sensors and *(iii)* "Depleted Field Effect Transistor" (DEPFET) sensors [125]. All three options are undergoing extensive research and development work. The CMOS Pixel based sensor MIMOSA-28 [124] was adapted and used at the vertex detector of the STAR experiment located at BNL [126].

The pixel sizes of, e.g., 17$\times$85 μm for the CMOS and 20$\times$20 μm for the DEPFET sensors enable the required point resolution.

The vertex detector's sensor material must perform while receiving an ionising dose of 1 kGy and a fluence (i.e., the number of incident particles per cross-sectional area) of $10^{11}\,\mathrm{n}_{eq}/\mathrm{cm}^2$ per year.

The largest background source for the vertex detector consists of the electron and positron pairs emitted via beamstrahlung. While most of the pairs remain inside the beam pipe, approximately 6 hits per cm^2 and bunch crossing are expected to reach the first layer. However, it is possible to identify and remove these hits since the pairs have a low transverse momentum.

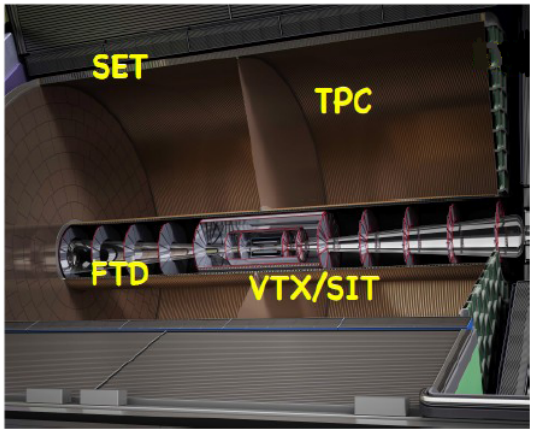


Figure 5.2: Sketch illustrating the layout and components of the ILD tracking system. Figure taken from [109].

5.1.2 The Silicon Tracking System

The silicon detectors are grouped in two categories: *(i) the central system*, located at the core of the barrel region, comprising the **SIT** (Silicon Internal Tracker) and the **SET** (Silicon External Tracker) while *(ii) the forward system* consists of the **FTD** (Forward Tracking Detector). The position of all the subcomponents are shown schematically in figure 5.2.
The detectors must comply with stringent requirements involving a low material budget and operation at very low power. The first condition can be fulfilled by minimising the detector support materials while the latter can be achieved by power pulsing. The research and development of the ILD silicon tracking system has been performed by the SiLC collaboration [127], [128].

The Central System

The SIT and SET not only contribute to improving the particle momentum measurements but also provide time-stamping information by correlating the hits in the silicon detectors to those in the TPC, thus enabling the correct bunch tagging of each event. They can also help in the alignment of the overall tracking. Furthermore, due to their location, the SIT and SET can provide monitoring of unwanted TPC distortions.

The internal barrel component, SIT, is situated radially in the gap between the vertex detector and the main tracker. By providing space points before the TPC, its main purpose is to provide a link between the vertex detector and the tracker.

The SET is positioned outside the main tracker (figure 5.2) and occupies the space between the TPC barrel and the inner surface of the electromagnetic calorimeter (ECAL). It provides a valuable entry point to the ECAL and also acts as a third silicon layer in the barrel region, thus improving the global momentum resolution.

The sub-detectors that make up the central system consist of layers constructed out of two single-sided sensitive strips which are tilted at a small angle with

respect to each other as detection devices. The SIT comprises two layers, while SET is made out of only one layer. The same type of sensor was chosen to be used in all sub-detectors.

The microstrip sensor that constitutes the basic detector unit for the SIT and SET has an area of $10 \times 10\,\mathrm{cm}^2$ with $200\,\mu\mathrm{m}$ thick silicon and a readout pitch of $50\,\mu\mathrm{m}$. The strip is virtually edgeless, with an inactive edge in the order of only a few tens of μm.

The Forward System

In order to perform flavour tagging on forward jets, it is essential that the trajectories of the charged particles contained in the jet can be extrapolated to the interaction point (IP).

The FTD is situated in the very forward and innermost detector region, covering down to an angle of $\approx 7^\circ$ with respect to the beam axis, thus, providing full tracking hermeticity to the ILD.

The forward tracker design must address the following challenges:

- **Reduced track bending:** The bending effect of the magnetic field becomes progresively smaller the closer the trajectory of a charged particle is to the detector z-axis. Consequently, the momentum measurement from the track becomes very challenging.

- **High occupancy:** Due to its position, the FTD will inevitably pick-up background events from the interaction region.

Consequently, the FTD design was optimised in order to accomodate these issues. It consists of seven disks positioned in the space between the beam pipe and the inner field cage of the TPC. The first disk is located at a distance of 220 mm from the IP, while the last one is placed at 2250 mm. This ensures a large lever arm which is crucial for momentum measurements since it can counteract the small track bending effect of the magnetic field.

The first two disks of the FTD are produced as pixel detectors that comply well with the high occupancy while the remaining five disks are made up of strip detectors. Similarly to the vertex detector, three different technologies are currently under consideration for the pixel disks: 50 μm thin CMOS based sensors, CCD and DEPFET. Their potential implementation is presented in detail in [123].

5.1.3 Time Projection Chamber

The task of the central tracker is to measure the trajectories of the charged particles produced in a collision event. Reconstructing these trajectories can provide crucial information regarding the transverse momenta and track parameters of the charged particles.

The Particle Flow paradigm places stringent requirements on the main tracker performance since the goal is to determine the energy of all the charged particles

entirely from its measurements. The (transverse) momentum resolution is typically parametrised as: $\sigma_{1/p_T} = a \oplus b/p_T sin\theta$, where a denotes the track curvature error and b describes the imprecision due to multiple scattering in the tracker's volume. In order to achieve the desired jet energy resolution, the objective for the tracking precision at ILD was set: parameter a should not exceed a value of $2\times10^{-5}\,\mathrm{GeV}^{-1}$, while the upper limit for b is 1×10^{-3}. This level of precision is two orders of magnitude better than the one previously achieved at LEP with the ALEPH detector, e.g., a_{ALEPH}=$1.2 \cdot 10^{-3}\,\mathrm{GeV}^{-1}$ [129].

At ILD, the main tracker is a 4.7 m long time projection chamber (TPC), centred around the beam axis, with an inner radius of ∼0.33 m and an outer radius of 1.8 m. The TPC is placed in a magnetic field of 3.5 T. The entire design and optimisation effort of the LCTPC collaboration [130] is aimed at achieving the desired precision.

The ILD choice of using a TPC as the main tracker is motivated by a number of advantages summarised below:

→ Small amount of material before calorimeters: For a good energy measurement, a particle should undergo only the minimum amount of energy loss in the tracker before it reaches the calorimeter system. Since the TPC is essentially a gas filled chamber, it is equivalent to only $\approx 0.15\mathrm{X}_0$ [123] and, thus, fulfils the Particle Flow performance requirement. Furthermore, the small amount of material also reduces the probability that the $\approx10^3$ beamstrahlung photons present in the detector during a bunch-crossing [66] give rise to low energy e^+e^- pairs (pair background).

→ Particle identification: The amount of energy deposited in the TPC is different for each type of (charged) particle that crosses its volume with the same momentum (figure 31.15 of reference [131]). Thus, by performing $\mathrm{d}E/\mathrm{d}x$ measurements one can obtain a handle on particle identification. While this is currently not used in the ILD standard reconstruction, the TPC's capability of providing information on the nature of charged particles, i.e., leptonic or hadronic, would present several benefits: it could enable the use of the correct mass hypothesis in the track fitting and it would reduce the track-cluster association errors in the Particle Flow reconstruction, e.g., an (identified) electron associated to a large cluster fragment in the hadronic calorimeter.

→ Large number of measurements per track: For the ILD design, the tracks left by the charged particles can be measured with approximately 200 three-dimensional (in z and the r-ϕ plane) space points per track. This plays an important role in achieving the required momentum resolution. Moreover, this feature of almost "continuous tracking" also means that the decays of long lived neutral particles to charged daughters - either SM (e.g., K^0) or SUSY predicted particles - that happen inside the TPC volume can be measured and reconstructed. Similarly, it also enables the identification and reconstruction of tracks that feature kinks in their trajectory. Furthermore, the large number of measured points

also makes it possible for particles that undergo backscattering at the calorimeter interface and, thus, re-enter the TPC, to be identified and measured.

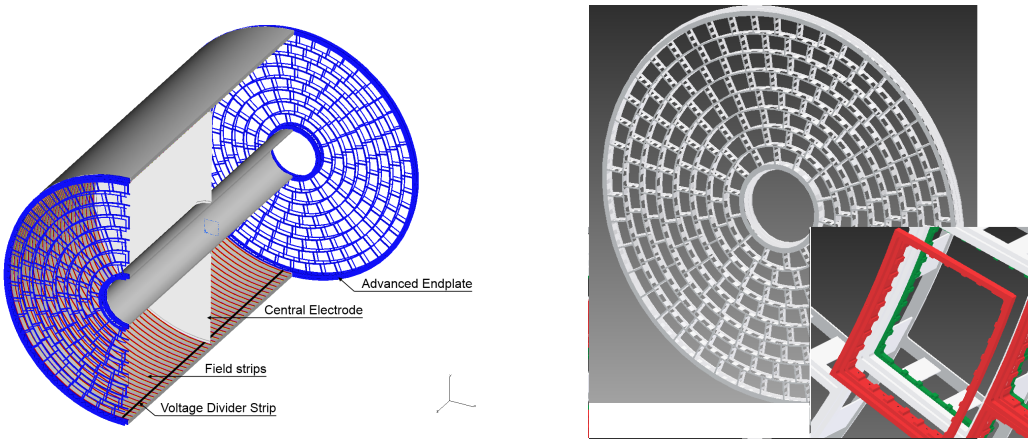


Figure 5.3: Conceptual design of the ILD TPC showing its main components (left) and a sketch of the endplate (right), including a detail of the support frame for the read-out modules. Both figures taken from [123].

The ILD TPC consists of a field cage that envelops two sensitive volumes separated by a central cathode plane. A schematic view of the ILD TPC design can be seen in figure 5.3. Each of the two extremities of the TPC constitutes an anode which also houses the read-out technology. The read-out pads are grouped in modules which will be concentrically assembled on the endplate as shown in the right panel of figure 5.3. Each module provides gas amplification, readout electronics, voltage supply and cooling. Currently, two different technologies for the gas amplification are under study: Micromegas [132] and Gas Electron Multipliers (GEM) [133].

The material of the endplate itself also plays an important role: if it is too thick it can lead to the production of secondary particles and, thus, hinder the Particle Flow reconstruction by increasing the number of association errors. Dedicated studies [134] have demonstrated that a material budget of up to $25\%X_0$ is acceptable and does not decrease the jet energy resolution dramatically.

In order to obtain a good momentum resolution, the diffusion of the electrons/ions drifting in the TPC volume must be minimal. This is highly dependent on the type of gas chosen to fill the TPC volume. For the ILD TPC, which has a drift length of $\approx$2 m, it was shown [135] that a gas mixture of Ar-CF4(3%)-isobutane(2%) known as the T2K gas [136] is worth taking into consideration.

Two potential sources of deterioration of the tracker's performance have been identified: *(i)* magnetic field distortions, *(ii)* the presence of ions in the TPC active volume. The former requires an active monitoring of the magnetic field during run time while the latter can be improved by introducing an ion gate at the anode.

5.2 Calorimetry

Calorimeters measure the amount of energy deposited in their instrumented volume. Under the Particle Flow paradigm, it is only photons and neutral hadrons that are to be reconstructed and measured in the calorimeters. However, the need to reduce the potential errors in clustering and track-cluster associations places further requirements on the calorimeters. Essentially, the Particle Flow concept relies on performing pattern recognition using the information provided by the calorimeters. Thus, they must act as imaging detectors that, by means of fine transverse and longitudinal segmentation, enable each particle to be "followed" throughout the detector. The high granularity of the calorimeters is crucial for distinguishing energy deposits produced by different particles.

The general layout of the ILD calorimeter system consists of a cylindrical-shaped barrel part and two end-caps. The system comprises an electromagnetic calorimeter (ECAL), as the innermost component, and a hadronic calorimeter (HCAL) which envelops the former. Both are installed inside the magnetic coil in order to reduce the amount of dead material in front of the calorimeters. In the forward region, a set of three calorimetric systems will be installed such that they: *(i)* act as a luminosity monitor (LCAL), *(ii)* measure the beamstrahlung (BeamCal) and *(iii)* provide hermeticity by detecting particles down to very low angles (LHCAL and BeamCal).

5.2.1 The ECAL

The mechanical design of the ECAL approximates the intended circular cylindrical shape with an octagonal prism. The half length of the barrel part is 2.35 m. Since the separation of energy deposits improves with larger ECAL inner radii [7], the ILD design foresees an inner radius of 1.84 m. Due to the compact design, the outer radius of the ILD electromagnetic calorimeter stops at 2.028 m.

The choice of absorber material must also fulfil the Particle Flow requirements. The aim is to enable the accurate separation between different calorimetric showers. Therefore, the following considerations regarding the material properties were taken into account:

- A **small Molière radius** (R_M) would be helpful for separating nearby showers.
- A **short radiation length** (X_0) would ensure that the electromagnetic showers start as early as possible in the ECAL.
- A **large interaction length** (λ_I) would mean that the number of *hadronic* showers that start in the ECAL would be reduced as much as possible.

Thus, the optimal material would provide a large λ_I/X_0 ratio. Moreover, since the calorimeters are placed within the magnetic coil, a more compact ECAL would greatly improve the overall cost-effectiveness by reducing the size of the coil.

Consequently, the absorber material chosen for the ILD electromagnetic calorimeter is tungsten which is dense, has a small Molière radius of R_M=9 mm and is characterised by a large interaction length to radiation length ratio: $\lambda_I/X_0 = 99/3.5$ [123].

The default baseline configuration of the ILD ECAL consists of a remarkable longitudinal segmentation into 30 sensitive layers which are all individually read. Thus, the Particle Flow reconstruction will benefit from up to 30 measurements carried out as the electromagnetic shower develops in depth. The sensitive layers are interspersed with 29 layers of tungsten absorber, which corresponds to $24X_0$. All this is achieved within a total thickness of less than 20 cm.

The transverse granularity of the ECAL is of utmost importance for minimising the potential association errors in the Particle Flow reconstruction. The ILD optimisation studies [109] demonstrated that an ECAL cell with the unprecedented size of $5\times5\,mm^2$ provides the best jet energy resolution. For the sensitive layers, two technologies were considered: *(i)* silicon pin diodes which are already available in the required size and provide straightforward operation and *(ii)* scintillator strips of $5\times45\,mm^2$, arranged in alternate directions in order to provide the desired transverse granularity. The latter option may unnecessarily complicate the pattern recognition procedure. The studies presented in this thesis are based on the silicon-tungsten (Si-W) option.

The performance of the ECAL design was studied both in simulation as well as in numerous test-beam campaigns. The CALICE collaboration [137] has designed and constructed prototypes for both material options. The obtained results are compatible with the energy resolution values assumed for the full ILD detector simulation. For instance, for the Si-W prototype described in [138], an energy resolution of $\Delta E/E = (16.6/\sqrt{E}) \oplus 1.1\%$ was obtained. The parameter values obtained from the fit to the simulated data agree within 5% to the measured resolution.

5.2.2 The HCAL

The ILD design foresees a cylindrical sampling hadronic calorimeter that uses steel as absorber material and can accommodate either scintillator tiles (analogue HCAL) or gaseous sensors (semi-digital HCAL) for the active part. The ILD HCAL has an inner radius of 2.058 m, while the outer radius is 3.410 m and the half-length of the barrel part is 2.350 m.

The ILD hadronic calorimeter is situated inside the magnetic coil. This positioning ensures the full compliance with the Particle Flow requirement that particles should traverse the smallest amount of material before reaching the calorimeters.

It is important that the ILD HCAL absorber material is non-magnetic and that it enables a compact and, hence, cost-effective design. Consequently, the chosen absorber material is stainless steel. The iron provides a satisfactory ratio between the interaction length and the electromagnetic radiation length: $\lambda_I/X_0 = 17\,cm/1.8\,cm$, thus, ensuring that the detector's volume is small. Furthermore, the

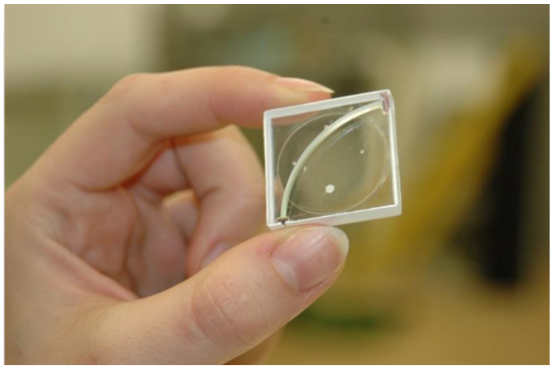


Figure 5.4: Picture illustrating the size, i.e., a surface area of $3{\times}3\,\mathrm{cm}^2$ and 3 mm thickness, of an ILD hadronic calorimeter tile. The image was taken from [139].

stainless steel is a rigid material that allows the construction of a self-supporting structure. This minimises the amount of uninstrumented space that would have been required for auxiliary support.

The work presented in this thesis is based on the analogue HCAL option. For this design, the sensitive layers consist of scintillator tiles which include silicon photo-diodes (SiPMs). They can measure the deposited energy as well as provide information on the position of the interaction. Furthermore, the response of the scintillator tiles is homogeneous. They are also only 3 mm thick which means that they are a good choice for a compact HCAL design.

Since the transverse granularity is essential for a good Particle Flow performance, the ILD optimisation studies [109] have shown that a size of $3{\times}3\,\mathrm{cm}^2$ for an HCAL tile provides the best jet energy resolution. Figure 5.4 shows the size and aspect of a typical ILD HCAL tile.

The mechanical design of the barrel part, forsees two options: *(i)* the segmentation of the HCAL barrel along the z-axis in two large rings, with each ring comprising 16 modules separated in the azimuthal direction or *(ii)* the division of the barrel into 5 rings along the z-axis with each ring being constructed from 8 modules. In both cases, each module would consist of 48 stainless steel absorber plates, with a thickness of 16 mm each, corresponding to λ_I=4.5 interaction lengths. The absorbers are interleaved with 48 scintillator layers, ensuring an unprecedented longitudinal granularity. Each sensitive layer is read-out individually, thus providing the Particle Flow reconstruction with up to 48 measurements for the hadronic showers.

The performance of the proposed ILD analogue HCAL design was studied by the CALICE collaboration in various test-beam campaigns. The energy resolution of the analogue HCAL physics protoype was measured [140] to be: $\sigma_{reco}/E_{reco} = 57.6 \pm 0.4/\sqrt{E_{beam}} \oplus 1.6 \pm 0.3 \oplus 0.18/E_{beam}$.

Applying software compensation techniques, the resolution is improved by almost 20%, reaching a value of $(45.8{\pm}0.3)\%/\sqrt{E}$ in the case of the stochastic term.

5.2.3 Forward Region Calorimeters

The forward region of the ILD calorimetric system is formed by three subdetectors: BeamCal, LumiCal and LHCAL. This section provides an overview of their main design features and expected level of performance.

LumiCal

The aim of the LumiCal is to measure the luminosity on a per mille level at a centre of mass energy of 500 GeV. The luminosity (L) measurement would be carried out using Bhabha scattering events: $L=N_{evt}/\sigma_B$, where N_{evt} denotes the number of events measured in a certain range of the polar angle and σ_B is the differential cross section for the same angular range. The cross section of the Bhabha scattering can be calculated from theory with a high level of precision [141].

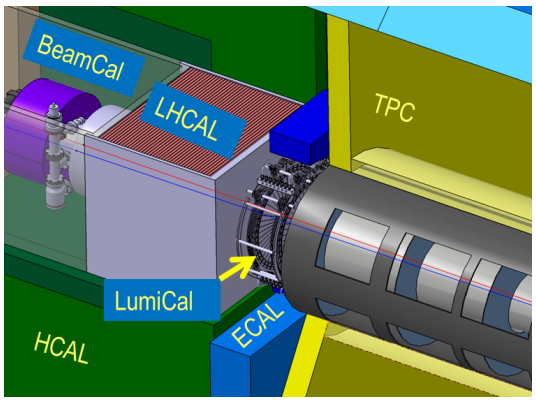


Figure 5.5: Sketch of the ILD forward region, illustrating the position of the forward calorimeters. Image taken from [123].

As shown in figure 5.5, LumiCal is situated in the circular opening of the ECAL endcap. It is centred around the beam and aligned with the electromagnetic calorimeter. The design consists of a cylindrical calorimeter with tungsten absorber disks, providing a total thickness equivalent to one radiation length, alternating with segmented silicon disks for the sensitive layers. LumiCal covers a polar angle range from 31 to 77 mrad (i.e., 1.8° to 4.4°).

The LumiCal performance, in terms of the energy resolution, was investigated by means of a simulation study presented in detail in [142]. The relative resolution $\sigma_E/E=a_{res}/\sqrt{E_{beam}}$ was determined to be $a_{res}=0.21\pm0.02\sqrt{GeV}$ for a beam of electrons with an energy of $E_{beam}=250$ GeV.

BeamCal

The BeamCal provides a fast luminosity measurement, on a bunch-by-bunch basis, and contributes to the beam tuning procedure. Furthermore, it can also measure the beam parameters with a precision of approximately 10% [123]. It is located immediately in front of the final focussing quadrupole magnet, as indicated in 5.5.

A large number of low energy electron-positron pairs produced by beamstrahlung would hit the BeamCal. It was estimated [123] that for a sensitive layer located at

an electron shower maximum and considering the nominal beam parameters the radiation dose would be 1 MGy after one year of operation. For this purpose, it is very important that the sensor technology that would be used for instrumenting the sensitive layers of BeamCal be radiation hard. The basic principle of the BeamCal technical design is rather similar to the previosuly described LumiCal layout: a "sandwich" electromagnetic calorimeter with tungsten absorber layers interpersed with GaAs sensors constituting the sensitive layers. Its design and location allows BeamCal to cover an angular range from 5 to 40 mrad (0.3° to 2.3°).

LHCAL

Finally, LHCAL is a hadronic calorimeter that would be located between BeamCal and LumiCal, thus, closing the gap between the latter and the main calorimeter, as shown in figure 5.5. LHCAL would extend the angular range of the main hadronic calorimeter such that it can reach the same angular region as LumiCal. This would be beneficial for enhancing the very forward particle reconstruction and identification.

5.3 Outer Region

The presence of a strong magnetic field enveloping the main subdetectors plays an important role in separating the charged particles from the neutrals. For this purpose, the outermost layer of ILD comprises the magnet system which is very similar in terms of the technical design to the one being used at the Compact Muon Solenoid (CMS) detector located at CERN [143]. The only fundamental difference to the CMS magnet is the dimensions of the ILD setup.

The ILD outer region consists of three main parts, as illustrated in figure 5.6. The first component is the *superconducting solenoid coil* which has a total length of 7.35 m and is constructed from three modules (each 2.45 m long) interconnected both mechanically and electrically. The maximum working temperature of the coil would be 4.5 K.

The second part is the *antiDiD (Detector-integrated Dipole)* which is situated at the outer radius of the solenoid. It provides a magnetic field that contributes to reducing the non-uniform pairs background (produced by beamstrahlung) in the vertex detector and in the silicon tracking system.

The last component is the iron yoke which consists of a barrel part and two endcaps that are constructed as dodecagons. The iron return yoke provides the main mechanical support for the detector. Furthermore, together with the calorimeters it is also acting as a radiation shield. A detailed account of the design concept for the ILD magnet system is given in [144].

The iron yoke is instrumented both in the barrel as well as in the endcap regions, acting simultaneously as a muon detector and a tail catcher system, measuring the energy of the showers that have leaked outside of the sensitive volume of the calorimeters.

In the barrel part, the design of the instrumented region consists of a sensitive layer placed before the yoke followed by 10 sensitive layers situated at 14 cm

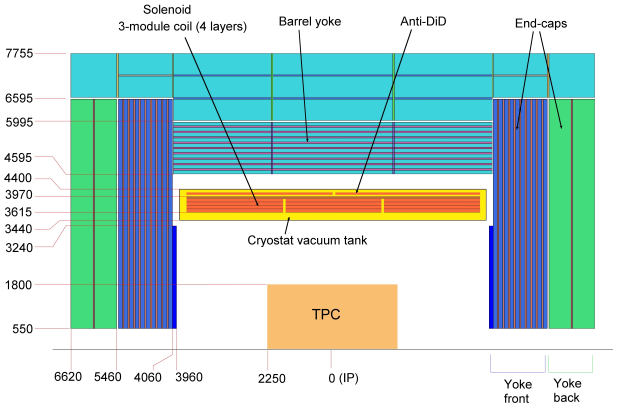


Figure 5.6: Sketch of the ILD outer region, illustrating the main components: solenoid coil, anti DiD and return yoke. Image taken from [123].

apart from each other and finally two sensitive layers separated by 60 cm. The same segmentation and distribution of the sensitive layers is used in the endcaps as well. The first 10 layers act as a calorimeter.

There are two different options taken into account for the sensor technology that would be used in the active layers: *(i)* scintillator strips with silicon photomultipliers as readout or *(ii)* resistive plate chambers [123].

Muon System

The ILD muon system can be reached by isolated muons with an energy above 3 GeV [109], while lower energy particles will be prevented from reaching it due to the strong magnetic field. The main task is to accurately match the signals recorded in the sensitive layers present in the return yoke to the energy deposits measured in the calorimeter (i.e., as a minimum ionising particle) and most importantly to the tracks recorded in the TPC.

Apart from the challenge posed by the multiple scattering that the isolated muons would suffer in the calorimeters and in the coil, another source of fake muons comes from pions.

A study of the identification efficiency of muons in jets as well as of the contamination with hadrons was performed with a full ILD detector simulation [123]. The instrumented ILD outer region was implemented in the detector simulation using the same design features as for the analogue HCAL active layers, i.e., $3\times3\,\mathrm{cm}^2$ tiles with SiPM readout.

The results demonstrate that for muons with an energy higher than 7 GeV, the identification efficiency is better than 97% even when the muons are not isolated, but part of a jet [123]. The wrongly identification of hadrons as muons lies at the level of a few percent. For less energetic muons, which either cannot reach the muon system because of interacting in the yoke material or are simply deflected in the magnetic field, the identification is made possible by the high granularity of the ILD calorimetric system.

Tail Catcher

As mentioned earlier, the first layers of the muon system also act as a tail catcher by measuring the energy of the calorimetric shower remnants that leak outside of the HCAL. A test study using muon and pion beams aimed at a prototype module of ILD was performed by the CALICE collaboration. The module included ECAL, HCAL and tail catcher/muon chamber (TCMT) prototypes. The experimental setup and the test beam instrumentation is described in detail in [145]. The investigation concluded that, for the standard ILD design, the TCMT would contribute by improving the relative energy resolution by 6 up to 16% for pions with an energy range between 20 and 80 GeV [145].

The International Large Detector was specifically designed and optimised for full compatibility with the Particle Flow paradigm. The ILD performance has been investigated using two subsequent implementations of a `GEANT-4` based full simulation. An overview of the ILD full simulation framework and the obtained detector performance are presented in the next chapter.

CHAPTER 6

The ILD Full Simulation and Reconstruction Framework

The design of the International Large Detector (ILD), summarised in chapter 5, was optimised to fulfil the Particle Flow requirements. A multitude of detector performance and optimisation studies have been carried out not only through test-beam campaigns, but also using a detailed detector simulation. Furthermore, all the International Linear Collider (ILC) physics studies, including the ones presented in this thesis, have been carried out by means of an ILD Monte Carlo simulation.

Collision events can be simulated either by using a thoroughly detailed description of the ILD, which is the realm of the so-called **full** simulation, or with a **fast** simulation that, in order to reduce the necessary computation time, uses some well motivated simplifications, both in the description as well as in the treatment of the detector response. This chapter describes the full ILD simulation and reconstruction chain.

As the ILD detector concept evolved, two major Monte Carlo data mass productions were carried out. Each of them consisted of: *(i)* simulated events corresponding to several chosen benchmark physics scenarios and *(ii)* a large number of Standard Model (SM) background simulated events.

The first data mass production was performed for the physics and optimisation studies discussed in the ILD "*Letter of Intent*" (`LoI`) document [109]. The second was produced with an even more detailed description of ILD and was used in carrying out studies for the ILD "*Detector Baseline Design*" (`DBD`) report [123]. As the detector design aspects evolved, the software implementation of the ILD Monte Carlo simulation also developed. The ILD simulation, reconstruction and analysis software framework under which both versions were developed is known as `iLCSoft` [146].

The present chapter provides an overview of the most important software tools in `iLCSoft` and their two subsequent implementations. The versions used for the *Letter of Intent* mass production will be referred to as the `LoI` versions, while the

ones employed in the *Detector Baseline Design* will be known as the DBD versions throughout this thesis. The detector performance results obtained with the two versions are also discussed and compared.

For consistency, all the ILD software tools rely on a standard event data model which is described in the following section.

6.1 The LCIO Event Data Model

For analysis as well as detector optimisation purposes, it is essential that, after each step in the data production chain, the key information is stored in a clear, easily accessible and consistent way. In order to serve this goal, the linear collider scientific community has developed an event data model (EDM) known as LCIO [147].

In LCIO, the information is stored on an event-by-event basis which is represented by the LCEvent class. This is the main data container which stores all the information related to one event. The event data is grouped into several LCCollections which are formed by a number of LCObjects of the same type. For example, an LCEvent will contain an LCCollection of tracks that have been reconstructed in the detector's tracking systems in that particular event, where each track is itself an LCObject. Metadata can be assigned to the LCEvent as well as to every LCCollection. For example, a jet collection may contain extra information for checking the quality parameters of a jet clustering algorithm.

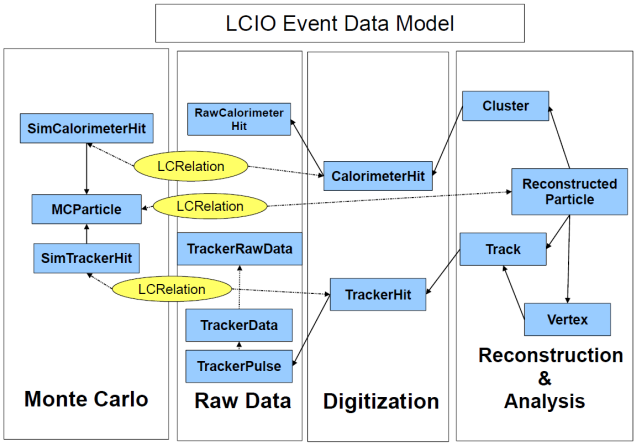


Figure 6.1: Overview of the key data classes in the LCIO event data model. Sketch taken from [148].

A schematic overview of the LCIO structure, featuring the main data classes and the relationship between them, is shown in figure 6.1. The schematic representation follows, from left to right, the order in which the data is processed in the simulation and reconstruction chain.

Thus, on the very left side, the chain begins with the Monte Carlo information for which the main class is the MCParticle. The MCParticle collection stores the *generator level* information about the Monte Carlo "true" particles. The detector

simulation can also produce MCParticles, as a result of the interactions with the detector material. These Monte Carlo particles, produced in simulation, are then added to the MCParticle collection.

The second panel presents the raw data classes which contain low level information as would be received by the subdetectors, e.g., channel numbers and uncalibrated electronic signal counts. This type of data undergoes a digitisation step that "translates" it into calorimeter and tracker hits.

The tracking algorithms use the digitised tracker hits to form reconstructed tracks. Finally, the Particle Flow reconstruction creates clusters from the digitised calorimeter hits and associates them to tracks. The result is a list of objects called ReconstructedParticles.

At the end of the data processing chain, the LCIO classes are stored in the standard slcio file format.

Each individual simulation step is summarised in the following sections.

6.2 Event Generation

The software tool chosen for event generation is WHIZARD [149]. Its general working principles are briefly described in the following.

WHIZARD is a program that calculates cross sections, obervables' distributions and generates events of hard scattering and decay processes that may occur at a high energy collider. The theoretical basis on which the program relies is leading-order perturbation theory.

The physics model to be considered in the event generation must be provided as input to WHIZARD, in the format prescribed by the SUSY Les Houches Accord (SLHA) [150]. WHIZARD allows flexibility in choosing the physics processes list. Thus, after a certain physics model, e.g., SM, MSSM, etc., has been selected by the user, a list of required reactions can be specified as input. The process complexity is only limited by the computing power, WHIZARD presently being able to compute scattering processes with up to eight particles in the final state. It can also simulate the radiation of photons (ISR) from the initial state electron and positron.

Furthermore, the program takes into account the luminosity weighted energy spectra of the initial state colliding particles (beamspectra) provided by the user. The energy spectra account both for beamstrahlung effects as well as for the intrinsic, machine induced, energy spread. For the ILD Monte Carlo mass productions, the beamspectra were created as two dimensional in order to accommodate a correlation between the two beams. They were calculated using the GuineaPig [151] program.

Once the processes have been specified, O'Mega, a program that generates matrix-elements for any reaction possible at tree level in the input physics model, is automatically called. O'Mega [152] is essentially a helicity amplitudes calculator that takes the Feynman rules and the relations between the coupling constants as input. Since the matrix elements are calculated as helicity amplitudes, this means that the spin and colour correlations are maintained.

After the matrix element calculation, WHIZARD performs the phase space integration. The showering and hadronisation were carried out within WHIZARD by calling PYTHIA [69].

Two different versions of WHIZARD were used in the mass productions: version 1.40 was used for the LoI while version 1.95 was employed for the DBD. While the software implementation evolved from one version to the other, e.g., in terms of bug and compatibility fixes, etc., there are no fundamental differences between the two WHIZARD versions. There were, however, several differences concerning the generation of the samples [123].

Firstly, as indicated in section 3.3, two different sets of beam parameter values were employed in the two subsequent ILD Monte Carlo mass productions. Thus, the LOI sample was generated using the so-called RDR beam parameters [55,109], while the DBD samples use the TDR beam parameters [57]. Both sets of parameter values are given in table 3.1.

In case of the RDR parameters, the beam energy spread at the interaction point (IP) is $\frac{\delta E}{E} = 0.14\,\%$ for electrons and $\frac{\delta E}{E} = 0.10\,\%$ for positrons. The TDR parameters lead to a smaller beam energy spread at the IP with $\frac{\delta E}{E} = 0.12\,\%$ for electrons and $\frac{\delta E}{E} = 0.07\,\%$ for positrons. These differences are perceptible when comparing the WHIZARD calculated cross-sections.

Another relevant difference concerns the PYTHIA parameters that determine the manner in which the program simulates the quark hadronisation. Thus, in case of the DBD Monte Carlo event generation, the OPAL parameters, described in section 2.2.1.2 of [123] and [153], were used. In contrast, the default PYTHIA parameters [154] were employed in the LoI mass production. This had the effect that the neutral, long-lived hadron multiplicities are over-estimated by up to 15% in the LoI Monte Carlo samples [123].

6.3 ILD Full Simulation

The main purpose of a full detector simulation is to output information that, for all intents and purposes, is as similar as possible to what the real detector would measure in a real collision event. Consequently, the ILD full simulation passes the WHIZARD generated physics events through the geometry of a very detailed detector description which includes not only the sensitive regions but also the service and support structures (e.g., cables, cooling, etc.). This enables a high degree of realism to be achieved. The detector response as well as the interactions of each particle (including secondary particles) are calculated from well established formulae and parametrisations of particle-matter interactions.

6.3.1 The Mokka Application

The ILD full simulation has been realised as the GEANT4 [155] based software package called Mokka [156]. All the ILD subdetectors have been implemented in Mokka with a high level of engineering detail.

For each ILD subdetector, e.g., the tracker (TPC), etc., its geometry description and component materials are stored in a MySQL [157] database. When an e^+e^- collision event is simulated, the whole ILD model is "constructed" from the previously defined subdetectors, by querying the MySQL database. The information stored in the database is transformed into a GEANT4 object by a C++ code known as a *driver*.

This modular approach has played a very important part in the detector optimisation work. The specific values stored in the database and the code that actually builds the detector are completely independent. Thus, while constantly checking for potential inconsistencies, the description (e.g., the dimensions) of any subdetector can be modified solely in the database, without having to modify the code of the driver. This important feature allows a high degree of flexibility in studying the potential benefits of changing various detector design aspects.

The input data for Mokka is a WHIZARD generator file, in the stdhep format which contains the basic physics information of the collision event. In Mokka, the Monte Carlo particles are passed through the ILD detector model where the interactions with the detector materials, including the formation of new Monte Carlo particles, are simulated. The output data follows the LCIO data format.

Two subsequent versions of Mokka were used for the major Monte Carlo data mass productions, i.e., v06-07-01 for the LoI and the even more detailed v08-00-02 for the DBD.

One of the tasks of this thesis was to investigate the impact of the increase in realism between the two versions of the full simulation on the Particle Flow performance, in the context of a physics study. Therefore, the most important differences between the LoI and the DBD versions are discussed in the following section.

6.3.2 Mokka Implementations: LoI and DBD

The first overall difference to be noted between the LoI and the DBD versions of the ILD full simulation concerns the treatment of the generator-level information about long lived and exotic particles. In the later Mokka version, the treatment of this information was improved, as mentioned in [123].

For the physics studies presented in this thesis the ILD_00 model from the LoI simulation version and the ILD_o1_v5 model from the DBD version are relevant. Their detailed description, including geometrical measurements, chosen materials and their segmentation, is presented in [158] and in tables III-1.1 and III-1.2 from [123], respectively. The differences between ILD_00 and ILD_o1_v5 most relevant for the present discussion will be addressed in the following.

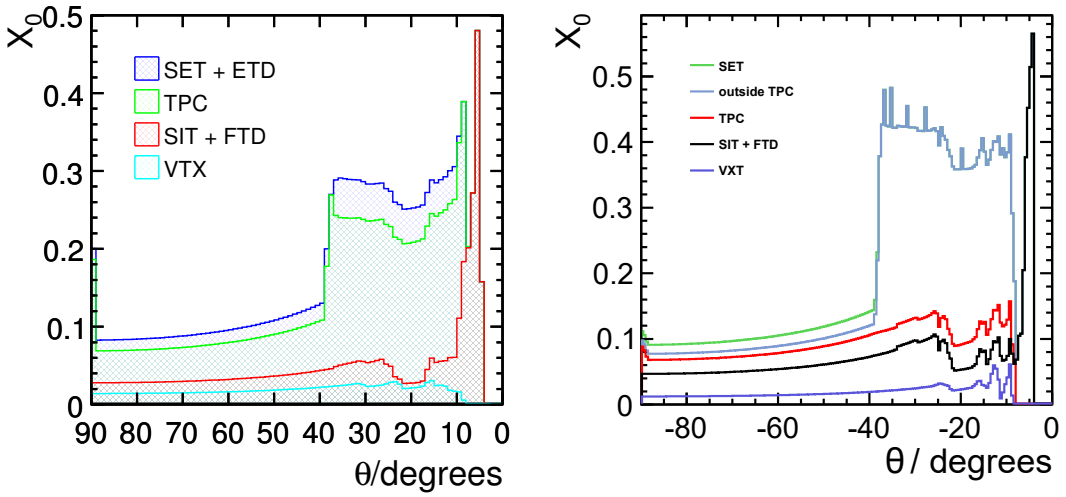
The DBD version of the ILD detector model represents an improvement in realism with respect to the LoI implementation which already contained a certain amount of support and service material. This was achieved in the DBD case by:

- further including gaps, imperfections and dead material in the detector description;
- adding material required for powering and cooling the subdetectors;

- detailing the description of the support structures, cabling and the electronics.

The implementation of all service and support materials is based on the estimates provided by the groups involved in the development of the ILD design. Furthermore, the description of some ILD subdetectors had to be completely rewritten for the DBD version since their design had not reached the required level of detail and realism at the time of the *Letter of Intent* document.

The differences between the ILD_00 and ILD_o1_v5 detector models can be observed in terms of the material budget which is presented in figure 6.2, in terms of the radiation length of the material as a function of the polar angle in the detector's coordinate system.



(a) LoI **(b)** DBD

Figure 6.2: The material budget for the ILD_00 model, (a), and for ILD_o1_v5, (b), expressed in terms of the integrated radiation lengths versus the detector polar angle, θ. In both cases, only the tracking detectors are considered. Plots taken from [109] and [123], respectively.

A general comparison of the two plots leads to the conclusion that the overall material budget has increased in the ILD_o1_v5 model, with respect to ILD_00 due to the increase in realism. As an example for added details, it can be seen from figure 6.2 that, at small polar angles, the DBD plot shows a set of spike-like features, in contrast to the LoI. These features correspond to the electronics, cabling and the required support structures located around the TPC endcap region that were added in the DBD detector model.

The small peak in the TPC line at approximately 90° (marked as -90° in the DBD plot) represents the small cathode membrane of the ILD main tracker. Its material budget has been reduced in the DBD case.

Finally, the most visible change between the two detector models comes from the implementation of the TPC endplate (marked as "outside TPC" in figure 6.2b) which, due to the improved realism, has increased in the DBD simulation by almost 50% in terms of material budget with respect to the LoI case.

The design of the hadronic and electromagnetic calorimeters has a strong impact on the Particle Flow performance. Therefore, in the context of this comparison, it is relevant to note that both `ILD_00` and `ILD_o1_v5` use the same materials, technology and the same segmentations for both types of calorimeters.

Once the passing of the generated Monte Carlo particles through the detector volume has been simulated the task of the full simulation is complete. The next step towards analysing the physics event is the reconstruction of what has occurred in the event.

Corresponding to the two subsequent `Mokka` versions, there are two distinct implementations of the event reconstruction framework. For consistency, the same labels, `LoI` and `DBD`, will be used to distinguish between them. An overview of the whole ILD reconstruction framework is given in the following section and, wherever applicable, the differences between the `LoI` and `DBD` implementations will be addressed.

6.4 The `Marlin` Reconstruction Framework

The event reconstruction is carried out using the modular `C++` application framework `Marlin` [159]. Each step in the reconstruction chain is written as an individual, dedicated module, called *processor*. Every processor reads in the relevant data collection from the standard `LCIO` `Mokka` output, performs the specific reconstruction task and outputs a collection of reconstructed objects.

The ILD's geometrical properties must be known to the algorithms performing the reconstruction. This is achieved by means of the `Gear` package [160] which ensures access to the detector description information at both reconstruction and analysis level.

In `Marlin`, the processors that are to be used, the order in which they must be run as well as the values of their input parameters can be easily specified via an `xml` *steering* file. All requested processors are executed one after the other on an event-by-event basis. It is possible for one processor to take as input a data collection created previously by an "upstream" processor.

The modularity of the `Marlin` framework makes it a very useful environment for developing and studying analysis and reconstruction tools. For instance, two distinct jet clustering algorithms, implemented as two different processors, could be run subsequently over the exact same simulated events and their output could then be directly compared.

The main stages in the reconstruction process are summarised in the following:

- **Hit digitisation**: The hits simulated in the tracking detectors are smeared according to a parametrisation of the point resolution determined by the relevant R&D groups. The calorimeter hits simulated in the electromagnetic and hadronic calorimeters are scaled with a certain factor according to the determined sampling fraction.

- **Track reconstruction**: The hits detected in the tracking system are processed by a set of complex topological pattern recognition algorithms that

combine hit clustering methods and Kalman filter (KalTest) extrapolations for adding new hits. The DBD version of the track reconstruction consists of a series of new algorithms implemented as C++ packages. These new software packages replace the old pattern recognition processors, based on LEP techniques and FORTRAN code, that were used in the LoI version.

- **Particle Flow reconstruction**: The Particle Flow algorithm produces a comprehensive list of ReconstructedParticles by subsequently: *(i)* using the list of reconstructed tracks provided as input, *(ii)* applying a set of quality cuts on the tracks, *(iii)* forming clusters from the input calorimeter hits and *(iv)* matching the tracks with the created clusters. The Particle Flow package used for the ILD event reconstruction is PandoraPFA [7] which is the most mature implementation of this type of algorithm available. There are two implementations of PandoraPFA, each associated to the corresponding ILD simulation version, i.e., the LoI one is called PandoraPFA while the DBD version is known as PandoraPFANew. They will be discussed in more detail in section 6.5.

- **Vertex finding** and **Jet flavour tagging**: The LCFIVertex package [161] performs: *(i)* vertex finding, based on the "ZVTop" algorithm [162], *(ii)* the identification of heavy flavour jets, using artificial neural networks, and *(iii)* the estimation of the jet charge. This package was available for both the LoI and the DBD versions of the reconstruction. However, for the DBD, a new processor called LCFIPlus was introduced with the aim of improving the flavour tagging. Consequently, the LCFIPlus package uses additional input variables, multi-variate analysis techniques and a new jet clustering algorithm.

- **Beam background overlay**: This step is performed *only* in the DBD version of the full reconstruction. The $\gamma\gamma \rightarrow$ hadrons (with low transverse momentum) background, presented in section 3.3.2, must be taken into account since it gives rise to real tracks and clusters. In the case of the ILC running at a centre-of-mass energy of 500 GeV, an average of 1.2 $\gamma\gamma$ events per bunch crossing are expected. It must be noted that, in the DBD Monte Carlo mass production, this number was inadvertently raised to 1.7 $\gamma\gamma$ events.

  Before performing the full DBD reconstruction, randomly chosen, fully simulated $\gamma\gamma \rightarrow$ hadrons events are overlaid, according to a Poisson statistic, on top of the main physics event. The z-position of the overlaid events is randomly smeared within a spread of 300 μm to reflect the fact that these events come from a different vertex than the one corresponding to the physics event. Both the main collision as well as the overlay events are "combined" and then processed together by the full reconstruction. Thereafter, it is the task of each physics analysis to establish the most effective procedure for distinguishing and removing the unwanted reconstructed particles

coming from the $\gamma\gamma \rightarrow$ hadrons overlay events from the ones belonging to the actual physics event.

The $\gamma\gamma \rightarrow$ hadrons overlay procedure is *not* performed for the `LoI` Monte Carlo samples.

Since the Particle Flow procedure is very important to the studies presented in this thesis, a more detailed overview of the existing algorithms in `PandoraPFA` as well as the differences between `PandoraPFA` and `PandoraPFANew` are highlighted in the next section.

6.5 The `PandoraPFA` Reconstruction Algorithm and Implementations

A Particle Flow algorithm carries out a sophisticated pattern recognition procedure in a complex environment, i.e., dealing with a large number of detector measurements. Its software implementations must handle the very high granularity of the ILD calorimeters and process the recorded data in a short amount of CPU time and with as little memory usage as possible. Furthermore, the pattern recognition should make few association errors, i.e., it should avoid double-counting the energy coming from the same particle and it should also accurately separate between energy deposits coming from different particles.

A software framework that performs the Particle Flow reconstruction and fulfils the software performance requirements described above was developed under the name of PandoraPFA [7]. Due to the especially high granularity of the ILD calorimeters, the PandoraPFA framework consists of more than 60 stand-alone algorithms.

The PandoraPFA implementation evolved in time, especially in terms of software engineering. The first implementation, `PandoraPFA` [7], was developed as a proof-of-principle. The Particle Flow performance evaluation presented in the ILD Letter of Intent [109] demonstrates not only the viability of the proposed ILD high-granularity design, but also that the Particle Flow principles and their implementation in `PandoraPFA` can indeed provide the jet energy resolution required for precision physics studies.

The proof-of-principle implementation of PandoraPFA was successful due to its highly sophisticated pattern recognition algorithms and their complex interplay. However, because of this complexity, it soon became evident that the further development of algorithms required a new central software framework that would be more flexible and that could better address memory management and data bookkeeping issues [163]. Furthermore, it was also necessary to add new algorithms, that would be better suited for handling the dense and boosted environments of high energy jets (i.e., above 250 GeV). Thus, in order to address these matters, the `PandoraPFA C++ Software Development Kit (SDK)` [164] framework was developed.

The algorithms used in `PandoraPFA` [7] were re-examined and re-implemented

in PandoraPFA SDK [163]. This (re-)implementation is known as PandoraPFANew. While there are significant software differences between the two versions of PandoraPFA, there are not many relevant algorithm differences, especially in the case of low energy jets (below 180 GeV). Furthermore, the PandoraPFANew implementation was also written such that it can be detector and reconstruction framework (e.g., Marlin in the case of ILD) independent.

The main input to both PandoraPFA versions comprises the set of previously reconstructed tracks and the list of digitised ECAL and HCAL hits.

In PandoraPFANew, the decoupling of the pattern recognition algorithms from any specific detector geometry can also be seen, for example, in the implementation of the calorimeter hits: they represent simple points in space and time with an associated energy (or intensity) measurement. Extra information such as the detector description, type of the calorimeter readout technology and the hit's relative location in the detector can, however, be added.

The output of the PandoraPFA reconstruction is the same in both cases: a list of clusters, i.e., sets of combined calorimeter hits and a list of "*Particle Flow Objects*" (PFOs), i.e., the reconstructed particles of the collision event. A detailed description of the reconstruction procedures in PandoraPFA is given in [7].

The reconstruction operations that constitute the core of both PandoraPFA and PandoraPFANew are presented in the following. Wherever applicable, the differences between the two versions will also be discussed.

Track Selection and Topology: The reconstructed tracks, provided as input, first undergo a quality selection based on information such as: the number of hits they have produced in the TPC and/or the forward tracker, the position of their last hit with respect to the ECAL inner surface, their momentum uncertainty and their curvature. A dedicated Marlin processor searches for specific topologies of the reconstructed tracks, such as V^0s (i.e., decays of neutral particles into a pair of charged particles), *kinks* (i.e., decays of a charged particles into another particle of the same charge plus a number of neutral particles, entailing a visible change of the trajectory's curvature) and *prongs* (i.e., τ decays). Once identified, a list of V^0 and kinks topologies are handed over to PandoraPFA.

Calorimeter Hit Selection and Ordering: The next stage consists of processing the list of digitised calorimeter hits. Each hit contains information on: *(i)* the (x, y, z) coordinates of its position in the detector coordinate system, *(ii)* the specific ECAL or HCAL layer where the hit was detected and *(iii)* the amount of energy deposited.

The hits are first calibrated, in a dedicated digitisation processor, according to whether they were registered in the ECAL or the HCAL. In the following step, additional geometry information, e.g., the size of the calorimeter cell that has recorded the hit, is added. Next, isolated deposits like the ones produced by low energy neutrons that can travel long distances from the point where they were produced are identified and removed. Finally, the remaining hits are ordered into so-called "pseudo-layers" that essentially mirror the general structure and layout

of the ILD calorimeters.

Clustering: At first, a dedicated photon identification algorithm that can be used exclusively on ECAL energy deposits is run. The aim is to first cluster and identify the ECAL clusters produced by photons.

Several improvements to the photon identification have been implemented in `PandoraPFANew` [165]. For instance, a set of selection cuts on ECAL clusters have been introduced in order to ensure their compatibility with the expected longitudinal and transverse profiles of electromagnetic showers [166].

Furthermore, two new algorithms have been introduced in `PandoraPFANew` with the aim of identifying and extracting clear muons and electrons before the rest of the reconstruction continues. Following these steps, the standard clustering algorithm is run over all the remaining calorimeter hits.

The default clustering algorithm is based on using the position of the track extrapolation to the ECAL inner surface for cluster seeding. In `PandoraPFANew`, however, other procedures that do not rely on this information have also been implemented. The clustering of the calorimeter hits is performed by a cone-based algorithm that proceeds from the first ECAL layer and continues outwards until it reaches the final HCAL layer [7].

Topological Cluster Merging: The clustering algorithm is designed such that it is more prone to splitting the energy deposits of a single particle into smaller cluster fragments rather than accidentally merging together hits from different particles into a single cluster. A series of algorithms that follow well-determined topological rules are then run to identify which previously produced clusters should be merged back together. The high granularity of the ILD calorimeters allows the merging algorithms to continue "following" the tracks further along the particles' paths through the calorimeters and merge cluster fragments with a high degree of accuracy.

Track-Cluster Associations: While track extrapolations can be used for cluster seeding, no track-cluster association is realised at that stage. Therefore, the reconstructed tracks must be matched to the formed clusters in a separate step. The procedure relies on comparing each cluster's properties, such as its position and direction, to every track's properties at the inner surface of the ECAL. The decision to make an association is based on linear and/or helicoidal fits to the track and cluster.

Re-clustering: If a significant discrepancy occurs between the track momentum and the energy of the cluster associated to it, the algorithm attempts to re-cluster the calorimetric energy deposits.

The *clustering* and *topological cluster merging* steps are re-run exclusively with the calorimeter hits that make up the discrepant cluster. In each iteration, the algorithms' parameters are modified such that the cone size becomes smaller. This

is done in order to split the original cluster until its energy becomes compatible with the track momentum. The most compatible solution is saved.

When no significant improvement can be found and the discrepancy between track and cluster remains, the original cluster is stored with its unaltered initial parameters as a neutral cluster.

Alternatively, the program also allows the option of an "Energy Flow"-like procedure that forces the cluster-track compatibility. In this case, the cluster formation is essentially "forced" by choosing energy deposits in calorimeter cells located along the track extrapolation into the calorimeter volume. The energy deposits are added along this trajectory until the track momentum is matched. The rest of the energy deposits, initially associated to the track, form neutral clusters.

Fragment Removal: Even after the cluster merging stage, there may still be a significant number of charged cluster fragments wrongly identified as produced by neutral particles. The aim of the fragment removal procedure is to identify these neutral fragments and merge them with the charged clusters they, in fact, belong to.

In order to do so, every neutral cluster (that was not identified as a photon) is compared to every charged cluster reconstructed in the event. Whether the two clusters can be merged depends on: *(i)* the clusters' location in the calorimeter with respect to each other and *(ii)* whether the track-cluster energy compatibility would improve, if the two clusters were merged.

Formation of Particle Flow Objects: Lastly, the information gathered in the previous steps concerning the reconstructed tracks, clusters and their associations is combined and the final list of reconstructed particles, termed Particle Flow Objects (PFOs) is created.

After a rudimentary particle identification procedure all the PFOs reconstructed in the event are stored, together with the associated information, in the `ReconstructedParticle` collection, following the `LCIO` standard.

The list of reconstructed particles provided by `PandoraPFA` offers a detailed description of what is most likely to have happened during the collision event.

6.6 ILD Performance Evaluation

The Particle Flow compliant design of the International Large Detector was presented in chapter 5 and in greater detail in reference [123]. The evaluation of the ILD global performance is crucial for understanding the behaviour of the detector as a whole and verifying that the precision goals can be met. These studies can be carried out only by means of a detector simulation. The remainder of this chapter focuses on discussing the performance studies performed with both the `LoI` and `DBD` (i.e., `PandoraPFA` and `PandoraPFANew`, respectively) versions of the full ILD simulation and reconstruction. The differences in performance between the two

implementations are also presented.

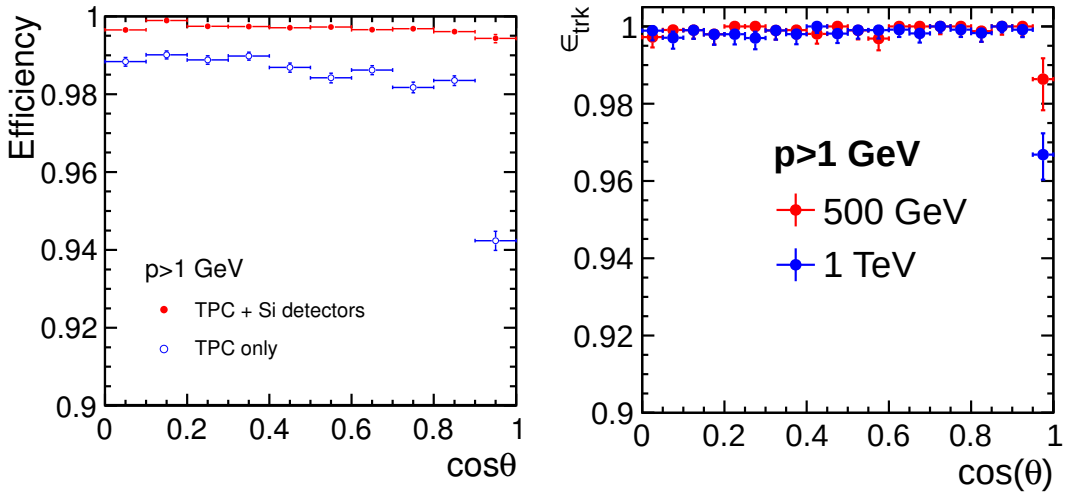
6.6.1 Performance of the Tracking System

The ILD tracking system benefits greatly from the interplay and complementarity between the TPC, as the main central tracker, and the silicon detectors' envelope. While the TPC provides continuous tracking (i.e., $\mathcal{O}(200)$ measurements per track) and hence a good input for pattern recognition, the vertexing detector (VTX) covers the very low p_T tracks that cannot reach the TPC and the forward tracker (FTD) ensures the hermeticity of ILD by reconstructing tracks down to a polar angle of $\approx 7°$.

The performance of the ILD tracking system is very important in view of the Particle Flow reconstruction. As presented in section 4.3, the energy of the charged particles would ideally be determined exclusively from the tracker measurements. The ILD tracking performance is discussed in the presenst section in terms of the track reconstruction efficiency and the transverse momentum resolution.

Tracking Efficiency

For both `LoI` and `DBD` versions, the track reconstruction efficiency was studied using Monte Carlo samples characterised by a high multiplicity in the final state, namely, $t\bar{t} \rightarrow 6$ jets processes. The results of these studies are presented in figure 6.3 where the reconstruction efficiency is expressed as a function of the cosine of the track polar angle. The left panel shows the track efficiency obtained with the `LoI` simulation and reconstruction, while the right panel illustrates the `DBD` result.



(a) `LoI` **(b)** `DBD`

Figure 6.3: The tracking efficiency as a function of the cosine of the track's polar angle. The left panel illustrates the `LoI` result and was taken from [109] while the right panel shows the result of the analogous `DBD` study and was taken from [123].

In both studies, the reconstructed tracks were compared only to those initial Monte Carlo tracks that: *(i)* start in a region of 10 cm in the vicinity of the interaction point, *(ii)* have a transverse momentum $p_T > 100\,\text{MeV}$ and *(iii)* a $\cos\theta < 0.99$. All the tracks produced by particles decaying in-flight, inside the tracking system, were excluded. In the DBD case, the pair background as well as the $\gamma\gamma \rightarrow$ hadrons background were taken into account.

In figure 6.3, the LoI (left) and DBD (right) results for the whole ILD tracking system can be compared following only the red markers. The data shown in blue is not relevant for this comparison.

Thus, in the DBD case, the average track reconstruction effiency is 99.7% [123] for all the tracks with a P_T higher than 1 GeV and over most of the polar angle range. This is true even for more forward tracks ($\theta \approx 18°$) as can be seen in the right panel of figure 6.3.

The LoI result, shown in the left panel, is well compatible with the DBD one. In the the LoI case, the track reconstruction efficiency for tracks with identical properties as mentioned above is 99.5% [109]. However, it can be observed that in the most forward bin the LoI-determined efficiency is significantly better than in the DBD case. This is most likely due to the addition of more material, i.e., improving the realism of the ILD simulation in the latter. Nevertheless, the LoI and the DBD results are very similar over most of the polar angle range.

Transverse Momentum Resolution

To estimate the transverse momentum resolution, both the LoI and DBD studies used fully simulated and reconstructed muon events in which the particles are produced at several fixed polar angles: 7°, 20°, 40° and 85°, over a whole set of momenta ranging from 1 to 200 GeV.

Figure 6.4, shows the transverse momentum resolution expressed as a function of the track momentum for different polar angles. The points represent the resolution values obtained from the analysis while the two continuous lines denote the resolution goals, as defined in 5.1.3, for two different values of the polar angle. The first one corresponds to more forward tracks (40°, in green) while the second describes rather central tracks (85°, in blue).

The LoI and DBD plots can be compared on the basis of the common colour code, while the marker style is not relevant.

In both cases, the resolution is visibly worse in the very forward region. This is due to the fact that at such shallow angles there is little difference between the orientation of the magnetic field and the direction of the track momentum. Nevertheless, the DBD performance at 7 ° is significantly better. When comparing the resolution goal at a polar angle of 85° (continuous blue line) to the corresponding obtained resolution values (markers), it is clearly visible that, in both studies, the momentum resolution requirement is fulfilled for the whole range of momenta larger than 1 GeV.

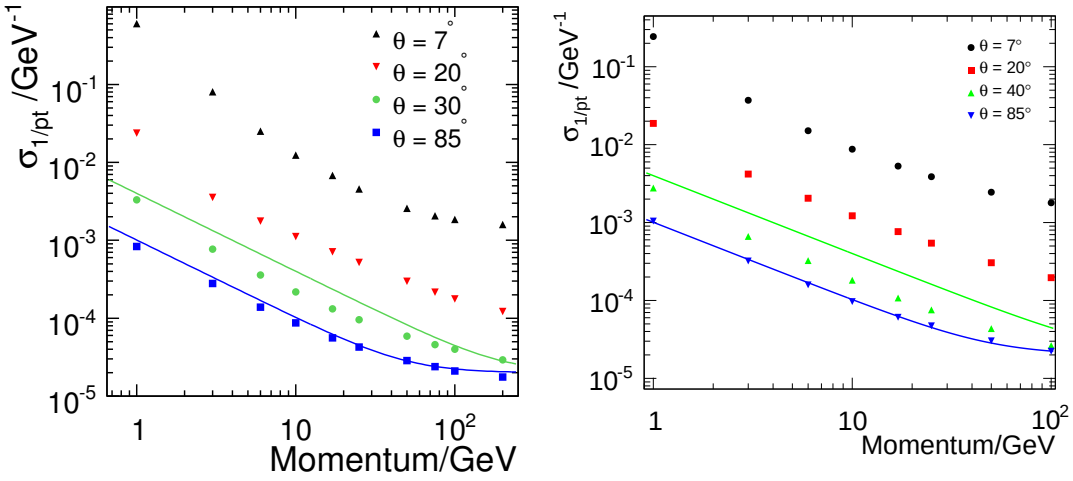


Figure 6.4: The p_T resolution as a function of the track momentum for four different track polar angles. The left panel illustrates the `LoI` result and was taken from [109] while the `DBD` result is shown on the right and was taken from [123]. Both plots have the same colour code.

6.6.2 Particle Flow Performance of ILD

Many physics scenarios to be studied at the ILC have final states with a high jet multiplicity. The Particle Flow performance of ILD is evaluated in terms of the *jet energy resolution* ($\sigma_{E_{jet}}$). The current section describes the methodology adopted by the ILC community for determining the jet energy resolution. The values obtained with the `LoI` and the `DBD` full ILD simulation and reconstruction versions are also presented.

Jet Energy Resolution Determination

The $\sigma_{E_{jet}}$ is typically evaluated using $Z \rightarrow q\bar{q}$ events. In these Monte Carlo samples, the Z boson is off-shell and produced at rest at the ILC. It then decays into a pair of quarks, producing two jets which are typically *mono-energetic* and *back-to-back*. Only the decays to light quark pairs ($u\bar{u}$, $d\bar{d}$, $s\bar{s}$) are considered. Thus, the potential contribution of the semileptonic heavy quark decays to the total jet energy resolution (σ_{HeavyQ} in equation 4.2) can be safely ignored. These samples will be referred to throughout this thesis as: $Z \rightarrow uds$ data.

For the $\sigma_{E_{jet}}$ evaluation, the relevant observable, obtained from the Particle Flow algorithm, is the total reconstructed visible energy in the event (E_{vis}). In view of this, one of the main advantages provided by the $Z \rightarrow uds$ samples is that the visible energy is clearly equivalent to the dijet energy ($E_{jj} \equiv E_{vis}$). Consequently, since the two jets are typically produced with equal amounts of energy, the single jet energy can be well approximated by $\mathrm{E_{jet}}=\sqrt{s}/2$. Thus, it is not necessary to run a dedicated jet clustering algorithm in order to distinguish between the two jets before determining their individual energy. The uncertainty on the jet energy resolution coming from the unavoidable jet clustering errors

($\sigma_{\text{Clust.}}$ in equation 4.2) can therefore also be ignored.

The jet energy resolution is defined in the following as:

$$\sigma_{E_{jet}} = \frac{rms(E_j)}{Mean(E_j)} \tag{6.1}$$

where rms denotes the root mean square. Since in the $Z \to uds$ events the jets are usually mono-energetic, it can be easily shown that the jet energy resolution, $\sigma_{E_{jet}}$, can be determined from the dijet energy (i.e., visible reconstructed energy) distribution using the following expression:

$$\sigma_{E_{jet}} = \frac{rms(E_j)}{Mean(E_j)} = \sqrt{2} \times \frac{rms(E_{jj})}{E_{jj}} \tag{6.2}$$

Thus, the average jet energy resolution can be determined from the total reconstructed visible energy. However, the distribution of the reconstructed energy is not necessarily Gaussian and, thus, the rms may not be the best estimator of the visible energy resolution.

This is clearly illustrated in figure 6.5. It shows the reconstructed energy, taken directly from the PandoraPFA output, for $Z \to q\bar{q}$ Monte Carlo events simulated at an ILC centre-of-mass energy of $\sqrt{s}$ =200 GeV. An angular cut was also applied such that the cases when the jets interact in the corner regions between the barrel and the endcaps are avoided [7].

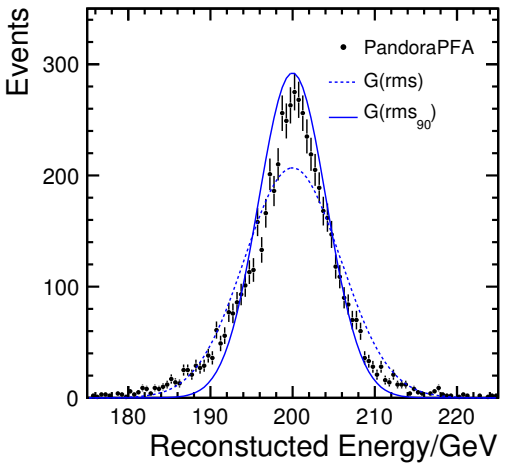


Figure 6.5: Reconstructed energy of $Z \to q\bar{q}$ simulated events. The dashed line depicts the best fit to the data using a Gaussian function. The continuous line illustrates a Gaussian distribution normalised to the same number of events and with a standard deviation equal to the newly defined rms_{90} estimator. Figure taken from [7].

It can be seen that the reconstructed energy distribution, illustrated by the black markers in figure 6.5, is *not* Gaussian. For comparison, the best result obtained from fitting a Gaussian function to the data distribution is depicted by the dotted line. The root mean square of the fitted Gaussian is 5.8 GeV [7].

The fact that the reconstructed energy does not follow a Gaussian distribution is not unexpected. The non-Gaussian tails are produced by those events for which the Particle Flow is hindered by confusion. Since the errors coming from mis-associations cannot be completely eliminated, the distribution of the reconstructed energy cannot be fully compatible with a Gaussian function.

Thus, the reconstructed energy resolution represented by the pure rms of the Gaussian fit, quoted above as 5.8 GeV, exaggerates the effects of the tails of the distribution. A different and more appropriate variable was adopted instead: the rms_{90} which is defined in [7] as the root mean square of the minimum region from the data distribution that contains 90% of the total number of entries. The rms_{90} is useful because it describes the resolution of the main core of the data, while remaining relatively unaffected by the tails of the distribution. In the case of the data presented in figure 6.5, the rms_{90} has a value of 4.1 GeV.

It must be mentioned that for a perfectly Gaussian-like distribution of the data, the rms_{90} is 21% smaller than the normal root mean square and thus it underestimates the energy resolution. However, as mentioned earlier, since the confusion during the reconstruction phase cannot be fully eliminated the distribution will never be perfectly Gaussian-like. Furthermore, in 6.5, the solid line depicts a Gaussian distribution, normalised to the same number of events, with a width σ=rms_{90}=4.1 GeV. It can be seen that, in the central region, it is 15% wider than the data distribution. This clearly demonstrates the non-Gaussian behaviour of the reconstructed energy distribution. Consequently, the jet energy resolution is re-defined as:

$$\boxed{\sigma_{E_{jet}} = \frac{rms_{90}(E_j)}{Mean_{90}(E_j)}} \tag{6.3}$$

where $Mean_{90}$ has an analogous definition as rms_{90}.

ILD Jet Energy Resolution Performance

In order to evaluate the jet energy resolution, a Monte Carlo study was performed using samples of $Z \to uds$ events, generated at four different centre-of-mass energies: 91.2 GeV, 180 GeV, 360 GeV and 500 GeV. The study was carried out both for the `LoI` [109] and the `DBD` [123] versions of the full ILD simulation and reconstruction software. Thus, `PandoraPFA` was run over the `LoI` data while `PandoraPFANew` was employed in the `DBD` case. Only jets that developed in the ILD barrel region (i.e., $|\cos(\theta_{q\bar{q}})| < 0.7$) were considered for this study.

The Particle Flow performance is presented in figure 6.6. The jet energy resolution, expressed as $\text{rms}_{90}(E_j)/Mean_{90}(E_j)$, is illustrated as a function of the individual jet energy.

It can be seen that, for both software versions, the single jet resolution is better than 3.7%, for jets with energies above 40 GeV, which fulfils the ILC precision requirements. However, in the `DBD` case, $\sigma_{E_{jet}}$ shows an improvement between 4% and 7% with respect to the corresponding `LoI` values. This is partially due to the improvements included in `PandoraPFANew`, especially regarding the treatment of high energy jets, as discussed in section 6.5. At the same time, the different parameter values used for the quark hadronisation in `PYTHIA` [69], as mentioned

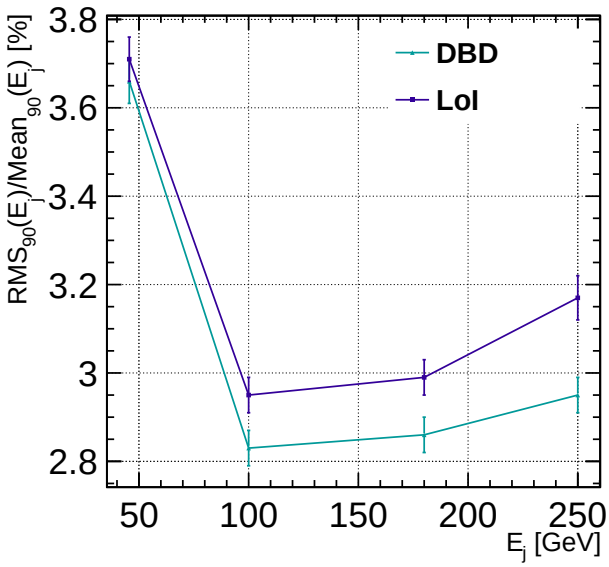


Figure 6.6: Comparison of the jet energy resolutions achieved with the `LoI` and the `DBD` versions of the ILD full simulation and PandoraPFA reconstruction. The figure was produced using the results presented in [109] for the `LoI` and in [123] for the `DBD`.

in section 6.2 also play a role: in the `LoI` Monte Carlo production, the amount of long-lived neutral hadrons is over-estimated. This is expected to lead to a worse determined jet energy resolution since: *(i)* the presence of more neutral hadrons in the final state carries a higher probability for Particle Flow confusion effects to occur and *(ii)* the energy of the neutral hadrons is typically measured with the least accuracy since, according to the Particle Flow principles, they are measured almost exclusively in the hadronic calorimeter.

Lastly, it must be noted that the increased realism of the detector description in the `DBD` version of `Mokka` was well compensated by significant software improvements such that it does not deteriorate the Particle Flow performance.

CHAPTER 7

SGV: The ILD Fast Simulation

The ILD full simulation, presented in the previous chapter, is characterised by a high degree of realism. This is achieved by using a comprehensive description of the detector's active and service materials and by taking into account the detailed geometrical description of the highly granular ILD design. However, this high level of detail requires a rather large amount of computing time. Presently, the field of particle physics is very dynamic, especially in view of the excellent performance of the LHC experiments. Thus, producing large amounts of Monte Carlo data in a short amount of time is highly desirable. This enables the ILC potential to be investigated with respect to any new physics that may be observed by other experiments.

The goals of this section are to introduce the concept of a *fast* detector simulation and to offer an overview of the various approaches and techniques presently implemented in different fast simulations programs. The main focus will be the detailed description of the ILD fast simulation - *La Simulation à Grande Vitesse* (SGV).

7.1 Motivation and Requirements for a Fast Simulation

While the role of the full detector simulation is very important both in physics as well as in detector optimisation studies, there are certain situations where its high level of detail can become a hindrance, especially in terms of computing time. This is the case, for example, in new particle searches, when performing model scans or in the study of rare processes. Similarly, the evaluation of the effects caused by backgrounds with large cross-sections also encounters such issues. In all these cases, a very large number of Monte Carlo events is required to provide the necessary statistics. Therefore, for this type of studies, the production of

Monte Carlo samples by means of the full simulation and reconstruction is no longer feasible.

The solution comes from employing a *fast* simulation which uses a simplified detector geometry description and most often a parametrised detector response. The simplifications used in the fast simulation clearly aim at reducing the necessary computing resources. However, the goal remains to provide Monte Carlo data that is as close as possible to what would be measured in real events and that accurately emulates the detector response.

As a concrete example, one of the main beam backgrounds at the International Linear Collider (ILC) is the reaction $\gamma\gamma \to hadrons$, in which two photons produced by beamstrahlung interact and give rise to real hadrons (described in section 3.3.2). The cross section of this process is large at approximately $\sigma_{\gamma\gamma} = 35{\times}10^3\,\text{pb}$ [6], as calculated with `PYTHIA` [154] for the ILC running at a centre-of-mass energy of $\sqrt{s}$=500 GeV. The first phase of the "H20" ILC operating scenario [81] foresees an ILC running time of about 4 years at $\sqrt{s}$=500 GeV which would correspond to $\int \mathcal{L}dt$=500 fb^{-1} of collected data. The expected number of $\gamma\gamma \to hadrons$ recorded events (N) can be estimated as $N{=}\sigma_{\gamma\gamma} \cdot \int \mathcal{L}dt$, which gives the very high number of $17.5{\times}10^9$ events.

The event generation of one $\gamma\gamma$ event, i.e., the step in the Monte Carlo production chain that precedes the passing of the outgoing particles through the detector simulation, takes approximately 10 ms [6]. However, the amount of time necessary to fully simulate and reconstruct such an event is in the order of a few minutes. This means that producing the whole Monte Carlo sample of this type of ILC beam background with the ILD full simulation and reconstruction would take approximately 3000 years of computing time. Even when considering the resources provided by a large computer cluster, the total CPU time needed extends over several years.

Furthermore, producing a Monte Carlo sample that has the same size as the real recorded data is not enough: the statistical errors of the simulation could affect (as systematics) the measured observables used in the analysis. In order to avoid this issue, it is usually required that the size of the simulated data set should be 5 to 10 times larger than the amount of real data. This requirement obviously increases the necessary computing time even more, such that the use of the full simulation and reconstruction for processes that have such large cross sections is no longer practical. This concrete example demonstrates not only the usefulness but also the necessity of developing and employing a fast detector simulation and reconstruction program.

The ultimate goal is to provide a fast simulation that emulates the detector response and the reconstruction output as much as possible. The specific requirements that a fast simulation for ILD must fullfil are summarised in the follwing:

- → The amount of time necessary for simulating one event should be of the same order of magnitude as the time required by the event generator to produce the event. In the case of the event generators typically used for producing ILC Monte Carlo samples, i.e., PYTHIA (version 6) [69] and WHIZARD (version 1.95) [167], the required time is on the order of $\mathcal{O}(10\,\text{ms})$. This is approximately $\mathcal{O}(10^3)$ faster than the full GEANT4-based simulation.
- → The detector description should be implemented in a flexible, easily adjustable way such that detector optimisation studies can be quickly carried out, if necessary.
- → The International Large Detector (ILD) concept is characterised by a very high granularity illustrated, for example, by the unprecedented ECAL cell size of $5{\times}5\,\text{mm}^2$ [123]. Consequently, a fast simulation of ILD must be able to simulate, in the shortest amount of time possible, a detector response that is compatible with this high level of granularity.
- → In order to achieve the realism required for physics studies, the ILD fast simulation must also take into account material effects.
- → The high granularity of the detector is a requirement for the Particle Flow reconstruction. Considering the precision necessary for physics analysis, the outcome of the ILD fast simulation should emulate as closely as possible the performance of the Particle Flow reconstruction, evaluated and studied in full simulation.

Various different strategies for the fast simulation of collision events have emerged over the years. The next section provides an overview of the different techniques available and motivates the choice of SGV as the fast simulation for ILD.

7.2 Overview of Fast Simulation Techniques

There are several fundamental working principles that most fast simulation programs rely on, for instance: *(i)* the simple smearing of the particles' four-momentum vectors, input directly from the event generator, assuming global values for the detector resolution and material effects, *(ii)* the simplification of the calorimeter simulation by using *frozen showers* or *(iii)* the use of covariance matrix calculating machines where the full covariance matrix is determined using information on the intial particles' four-momentum vectors and the detector design.

In the following section, all the methods mentioned above are briefly presented, following their current implementations and use either at the LHC experiments or at the ILC. The different techniques are compared and a special emphasis is placed on SGV and the motivation for designating it as the standard ILD fast simulation.

The following fast simulation software packages were considered for comparison:

- ❑ **The CMS Fast Simulation:** FASTSIM. The software framework of the CMS experiment comprises both a detailed full detector simulation based

on GEANT4 [155] and a fast simulation, FASTSIM [168], which is an object-oriented C++ based software package constantly validated and tuned with respect to the output of the full simulation.

- **ATLAS GEANT4 with Frozen Showers**. The ATLAS Collaboration has developed a complex software infrastructure for the Monte Carlo data production called the Integrated Simulation Framework [169] (ISF). In ISF, it is possible to use a full GEANT4-based tracker simulation combined with a *frozen showers* technique for the calorimetric system.
- DELPHES3. The DELPHES3 [170] package is a C++ based fast simulation of a generic collider experiment. It is capable of simulating a tracking system installed in a magnetic field, calorimeters and a muon system.
- **SGV: La Simulation à Grande Vitesse**. The SGV fast simulation program [6] was developed with the intial purpose of evaluating the upgrade of the DELPHI detector [171]. SGV has constantly evolved in the meantime and it has been intensely used for physics and detector optimisation studies not only at the DELPHI experiment, but also by the TESLA [172] project, one of the pre-cursor projects of the current International Linear Collider effort. While the further development of SGV is constantly on-going, the latest major revision and upgrade of the code itself comprised, among other improvements, the transition from a Fortran77 based code to Fortran95 plus the removal of certain library dependencies [6]. This led to the creation of the SGV 3-series that constitute the focus of this section.

The typical data used as input by a fast simulation program is a list of particles emerging from the collision event created by the event generator. The initial four-momentum and vertex of origin information is taken into account. In most cases, the generated particles are also allowed to decay while they traverse the detector simulation, according to their specific branching ratios and kinematics. The decay products are in turn propagated further through the fast simulation.

The techniques used for simulating the tracking system are rather different than the ones employed for the calorimeter simulation. Therefore, they will be discussed separately in the following comparison. The overview of the different fast simulation programs follows the discussions in references [173], [174] and [168] for the CMS FASTSIM, [169], [175] and [176] in the case of the ATLAS frozen showers technique, [170] and [177] for DELPHES3, [6] and [178] regarding SGV.

7.2.1 Fast Simulation of the Tracking System

The Particle Flow paradigm places strong requirements on the tracking system: the energy of most charged particles would ideally be determined from the tracker measurements. In the ILD design, the main tracker is a time projection chamber [123] which is complemented by an envelope of silicon trackers.

In terms of the tracker simulation, the CMS (FASTSIM) and the ATLAS (FATRAS [179]) fast simulation programs follow similar principles. Consequently, they will be exemplified by the CMS FASTSIM software package in the following.

The fast simulation programs mentioned previously will be compared in terms of: *(i)* the working principles of the tracking simulation, *(ii)* the implementation of the detector description and *(iii)* the types of particle interactions taken into account.

Working Principles

`FASTSIM`
The fundamental working principle of the CMS `FASTSIM` is that the material effects are described by *parametrisations at the hit level*. The outcome of an interaction between an incoming particle and the detector material is simulated by smearing the parameters of the hits (e.g., the position of the hit) produced by the particle, according to a previously established parametrisation function. The hits are then treated as input to the more involved, higher level reconstruction algorithms that are used in the full simulation as well [168].

A *simulated hit* is created at each intersection point between the particle's trajectory and a detector layer. Each simulated hit is then transformed into a *reconstructed hit*, by taking into account a finding efficiency parametrised with respect to the behaviour of the full simulation. The position of the reconstructed hit is obtained by applying a Gaussian smearing of the position of the simulated hit. The Gaussian resolution is parametrised as a function of relevant observables (e.g., the incident angle of the particle with respect to the tracker layer) and the parameter values are determined from studying the full simulation. Finally, the reconstructed hits are used in a fast tracking algorithm, described in [174].

`DELPHES3`
In `DELPHES3`, the propagation of a particle through the tracker volume consists solely of calculating its helix (or straight line) trajectory. The probability to reconstruct a track is parametrised as a function of the particle's transverse momentum (P_T) and pseudorapidity (η). The values of the parameters must be input by the user. Similarly, the momentum resolution must be user specified and it depends on the particle type (read from the Monte Carlo truth information), the value of its P_T and pseudorapidity. For each track the positions at the vertex and at the entry point in the calorimetric system are stored. The simulation does not take into account fake tracks, photon conversions and no $d\mathrm{E}/d\mathrm{x}$ measurements are performed [177].

`SGV`
The CMS tracker consists of three cylindrical layers of silicon pixels and ten layers of silicon strips [180]. Hence, the number of hits measured per track is an order of magnitude smaller than the expected ≈200 measurements to be performed for a track at ILD. Thus, considering the requirement to reduce the computing time as much as possible, it becomes obvious that the `FASTSIM` approach of simulating the material effects from parametrisations at hit level is no longer feasible in a fast simulation of the ILD TPC. A more efficient solution is provided by `SGV`.

In SGV, the simulation of a charged particle's trajectory in a magnetic field and the track reconstruction are essentially carried out in the same step.

The particle's helicoidal track is described by five geometrical parameters. The values of these parameters at the perigee[1] are calculated from the particle's Monte Carlo truth information. To emulate the tracker response, these initial parameter values are smeared using the corresponding entries from the covariance matrix of the track parameters computed at the perigee.

For the computation of the covariance matrix, SGV first determines how many measurements will be performed along the track and at which positions in the detector's coordinate system. Based on this information, SGV then uses a *Kalman filter* [181] approach to obtain the covariance matrix at the perigee. Since effects like, e.g., energy loss and multiple scattering, are taken into account in the calculation, no additional track reconstruction step is required.

Detector Description

FASTSIM

The technology used by the CMS tracker is different than the one of ILD, i.e., it consists of more than 15000 silicon strip modules. The FASTSIM *tracker simulation* uses a simplified description of the CMS tracking system employed by the full simulation. This abridged description [174] consists of more than 30 concentric cylinders that represent the sensitive layers. The material of the cylinders is pure silicon and it is assumed to be uniformly distributed. The thickness of each layer was tuned to the full simulation such that the number of bremsstrahlung photons is reproduced. The detector description does include non-instrumented cylinders that represent cables and service materials.

DELPHES3

The DELPHES3 program is not tuned by default to any particular detector geometry. To enable the straightforward customisation of the tracker design, its implementation is rather idealised. The tracking system is considered a uniform volume with neither service materials, e.g., the cathode or endplate read-out electronics, in the case of a TPC, nor gaps, for instance between read-out modules. These simplifications do not fullfil the realism requirements for an ILD fast simulation.

SGV

The SGV tracker description is a more generalised and flexible implementation of the same concept employed in FASTSIM. The detector can be described by a set of cylinders with a common axis, i.e., the z-axis, and a number of planes arranged perpendicularly to the common axis [178]. Each cylinder is determined by its radius and the minimum and maximum length in z. Analogously, a plane is described by its location on the z-axis and by its minimum and maximum radius.

[1] The *perigee* is the point on the track's projection in the x-y plane that is closest to the interaction point (i.e., the origin of the coordinate system).

The thickness of a cylinder or of a plane, in terms of radiation lengths, can be input by the user. This description enables a straightforward implementation not only of a silicon tracker, as in the case of CMS, but also of the ILD time projection chamber.

The granularity of the TPC read-out modules and, implicitly, the discrete number of measurements per track can be emulated by the presence of cylindrical surfaces in the SGV tracker description. The working principle will be described later in this chapter.

Introducing more realism, in SGV, the planes that describe the TPC endplates can be divided into active and dead material areas such that the non-instrumented boundaries between the TPC read-out modules can also be simulated. The flexibility and level of realism provided by the SGV detector description fulfil the requirements for the ILD tracker fast simulation.

Simulated Interactions

FASTSIM

In accordance with the technology used for the CMS tracking system, five interactions are simulated by FASTSIM in the tracker region: *(i)* electron bremsstrahlung, *(ii)* photon conversion, *(iii)* energy loss (in the case of charged particles) by means of ionisation, *(iv)* multiple scattering and *(v)* nuclear interactions. Their implementation is discussed in more detail in [168]. The first four are computed analytically from well-established formulae, as given in [99], while for the latter an event is randomly chosen from a sample of 2.5×10^6 nuclear interaction Monte Carlo events previously simulated in GEANT4 and overlaid in the fast simulation.

DELPHES3

The track trajectory calculation is purely mathematical in DELPHES3. There are no material effects taken into account at this stage and no smearing of the track parameters is performed. The DELPHES3 simulation assumes a perfect angular resolution. The propagation of the particles through the tracker is parametrised based on the magnitude of the magnetic field, the radius of the tracker and its half length - information which must be input into the simulation. Furthermore, the magnetic field, parallel to the beam axis, is considered to act only in the tracker volume. Thus, particles that have their origin outside of the tracker volume are ignored. This level of simplification is not compatible with the realism required by an ILD tracker fast simulation.

SGV

Similarly to FASTSIM, all the fundamental processes that typically take place in the tracking detectors can be simulated by SGV: photon conversion into an electron-positron pair, electron bremsstrahlung, multiple scattering and energy loss via ionisation [178]. The information regarding which interactions should occur, the energy threshold above which the particles can undergo the respective interactions and the intrinsic detector resolution must be provided by the user.

7.2.2 Fast Simulation of the Calorimetric System

Simulating the shower development in the calorimeters takes up a large fraction of the computing time in a full simulation program. In order to minimise this, most fast simulation packages rely on one of the following basic principles: *(i)* the shower development is emulated using a high degree of simplfication or *(ii)* the calorimeter response is simulated by smearing the particles' true energy with a user defined detector resolution. The first option is exemplified in this section by the CMS `FASTSIM` and the ATLAS `GEANT4` with frozen showers implementations.

The second technique can be illustrated by the `DELPHES3` program and `SGV`. In `DELPHES3`, the energy resolutions of the ECAL and of the HCAL are separately and independently parametrised as a function of the particle's pseudorapidity and energy. However, the calorimeter simulation is based on a number of assumptions and approximations [170] that are incompatible with the detector design and required level of realism for ILD: *(i)* the calorimetric system is outside the central B field, *(ii)* the segmentation of the calorimeters is regarded as uniform in ϕ and the same granularity is assumed for the electromagnetic (ECAL) and the hadronic calorimeters (HCAL), *(iii)* electrons (positrons) and photons are assumed to leave all their energy in the ECAL while the charged and neutral hadrons deposit all their energy exclusively in the HCAL *(iv)* the muons do not interact at all in the calorimetric system. Consequently, the second working principle of a fast calorimeter simulation will be exemplified in the following only by `SGV`.

`FASTSIM`

The CMS ECAL is essentially built from a large number of adjacent crystals. Therefore, its simplified description in `FASTSIM` is a homogeneous medium filled with lead tungstate crystal. The CMS hadronic calorimeter consists of layers of steel interspaced with plastic scintillator tiles.

In `FASTSIM`, the charged and neutral hadrons are propagated through the tracker and the calorimetric system. If the hadrons start showering in the tracker they are then replaced with the daughter particles which are then propagated independently. Otherwise, the hadronic showers may start either in the ECAL or in the HCAL.

The shower development in the ECAL is simulated using the Grindhammer parametrisation [182], [183], originally developed for the H1 experiment [95] at DESY. A calorimeter shower is regarded as a collection of energy deposits that need to be "placed" in the detector material. The energy deposits are spatially distributed according to the Grindhammer radial function. The azimuthal energy distribution is considered uniform in ϕ [173]. Other effects like shower leakage, energy loss occurring due to the gaps between the calorimeter modules and shower enlargement under the effect of the magnetic field are also taken into account.

In `FASTSIM`, the energy response of the HCAL is described by a double-sided Crystal Ball function, i.e., a Gaussian convoluted with a power law function to model the tail of the distribution. The parameters of the Crystal Ball function are determined from fully simulated samples of single charged pions with a transverse

momentum range between 2 and 300 GeV. This energy response is then applied for all simulated hadrons: for each hadronic shower the energy deposits are smeared using the corresponding Crystal Ball parameters obtained for that particular set of P_T and pseudorapidity (η) values. The energy deposits are spatially distributed in the HCAL materials using a parametrisation similar to the Grindhammer one.

The muons are propagated until they reach the muon chambers and they can undergo multiple scattering and energy loss via ionisation along their path. The calorimeters' response to muons is simulated in the same way as for pions.

`ATLAS frozen showers`

The ATLAS fast simulation framework uses another procedure for the calorimeter shower development. It is called "*frozen showers*" [176] and it is motivated by the observation that up to ≈80% of the computing time is usually taken up by the shower development of particles interacting in the electromagnetic calorimeter (ECAL). The frozen showers technique is used mainly for low energy particles that give rise to large cascades of secondaries in the ECAL. For highly energetic showers, i.e., that also reach the HCAL, a parametrisation approach that describes the shower development by means of a longitudinal and a radial distribution is employed (`FastCaloSim` [175]).

The frozen showers method consists of interrupting the shower development once the energy of the initiating particle is low enough to fulfil the halting condition and replacing it with a previously simulated (*frozen*) shower stored in a dedicated library, in the respective calorimeter region.

For this purpose, a shower library was first created using a full `GEANT4` detector simulation of the ATLAS calorimetric system. In this full simulation, low energy particles are started right in front of the calorimeter interface, at various positions and with different energy values. The spatial, timing and energy information of each individual hit that the particle produces as it passes through the detector are then stored. The frozen shower library is binned ("catalogued") in energy and position, the latter being expressed in terms of pseudorapidity.

In the frozen showers technique, the full `GEANT4`-based simulation is used to propagate the particles through the inner detector until they reach the calorimeters. The information regarding the type and energy of each individual particle is verified with respect to the shower stopping conditions. If they are fulfilled, a shower from the library is randomly chosen from the appropriate bin and is then overlayed in the ATLAS ECAL. Both the pseudorapidity and the energy bins are chosen randomly from the two adjacent bins, using certain probability distributions. After the appropriate shower is chosen, the total deposited energy is scaled to match the energy of the particle that is being replaced by the shower.

`SGV`

There is a significant difference in the design philosophy of the ATLAS and CMS calorimeters in comparison to ILD. The fundamental principle of the ILD design relies on very high transverse and depth granularity, required by the Particle Flow reconstruction.

For example, the CMS electromagnetic calorimeter [48] is constructed from 23 cm long $PbWO_4$ crystals with a transverse area of $\approx 2.6 \times 2.6\,cm^2$ in the barrel region. In contrast, the ILD ECAL has a depth of 18.5 cm in the same detector region and consists of 30 silicon sensor layers interspaced with tungsten absorber. Furthermore, the transverse granularity is significantly higher since the sensitive material is divided in cells with a $5 \times 5\,mm^2$ area.

This demonstrates that the shower development techniques used in the `FASTSIM` and ATLAS frozen showers are not feasible for an ILD fast simulation. Due to the high granularity, the necessary computing time would significantly exceed the required $\mathcal{O}(10\,ms)$ per event. Thus, in contrast to `FASTSIM` and ATLAS frozen showers, `SGV` does not produce a detailed simulation of the calorimetric showers' development. It directly emulates the outcome of a cluster reconstruction algorithm instead.

The calorimeters are described similarly to the tracker implementation, i.e., they comprise a series of nested cylinders for the barrel part and planes for the endcaps. In order to simulate a highly granular detector, each cylinder or plane can be divided into adjacent sensitive or non-measuring sectors.

The simulation first establishes, for each particle, whether it reaches and subsequently interacts in the calorimeters. If this is the case, the type of interaction, i.e., giving rise to a hadronic or electromagnetic shower, is also selected. The energy of the shower is obtained by randomising the particle's Monte Carlo truth energy, taking into account a user defined ECAL or HCAL resolution, respectively, and a reconstruction efficiency. Thus, `SGV` outputs the equivalent of a shower reconstruction procedure. This approach has the advantage that the time consuming task of simulating calorimetric showers is avoided.

The four programs considered for illustrating the various fundamental working principles used in fast simulations can be also compared in terms of their CPU time performance and complexity. This comparison is presented in the following section.

7.2.3 Performance

In order to meet the processing time requirements, the fast simulation programs must emulate not only the detector response, but also the output of the reconstruction stage. Table 7.1 summarises the various fast simulation techniques discussed in the previous section.

The CMS `FASTSIM` and ATLAS `frozen showers` implementations both require an additional reconstruction step. In `FASTSIM`, the tracker response is parametrised at hit level while the calorimeter shower development is described by a set of functions with parameters adjusted to the full simulation. The `frozen showers` method relies on a full `GEANT4` simulation of the tracker and randomly samples a large library of calorimetric showers. Thus, in both cases, the output of the fast simulation consists of a set of tracker hits for the charged particles and energy deposits in the calorimeters for both neutral and charged particles. This data must

	Frozen showers	FASTSIM	DELPHES3	SGV
Software framework	C++ based	C++ based	C++ based	Fortran 95
Tracker simulation	GEANT4 full sim.	parametrisations at hit level	geometrical track extrapolations	covariance matrix computations
Calorim. simulation	GEANT4 shower library	Grindhammer & Crystal Ball parametrisations	Log-normal param.	Gaussian param.
Further recon. required	Yes	Yes	No	No
CPU time	$\mathcal{O}(10^2)$s	4 s	7-9 ms	10 ms

Table 7.1: Comparison of the major fast simulation programs presently available, in terms of working principles and necessary computing time. The quoted CPU time estimates refer to a single simulated and "reconstructed" event for all four programs.

undergo an extra reconstruction stage before it can be used in physics analysis.

The CMS and ATLAS software frameworks allow the information provided by their respective fast simulations to be processed with the same software tools used in the full reconstruction. However, that is clearly not time effective. Consequently, both collaborations have developed fast digitisation and reconstruction plug-ins.

For instance, in the case of FASTSIM, a fast track reconstruction algorithm is used. It does not perform any pattern recognition but simply fits the tracker hits known to belong to a simulated track. The method uses the same fitting algorithms that are implemented in the full reconstruction. The track seeding efficiency is also emulated and, as a simplification, the hits that contribute to a too large χ^2 of the track fit are removed.

The FASTSIM performance in terms of computing time is remarkable. The simulation and reconstruction time required for a $t\bar{t}$ event generated at an LHC centre-of-mass energy of $\sqrt{s}$=13 TeV is 4 seconds [168].

The ATLAS full simulation requires an amount of CPU time of the order of $\mathcal{O}(10^3)$s [169]. The frozen showers technique reduces the processing time by a factor of two for a typical QCD event [176].

The DELPHES3 fast simulation directly outputs reconstructed photons and charged leptons. The photons, muons and electrons are identified by making use of the Monte Carlo truth in the form of their PDG code [177]. The step that transforms the simulated information into reconstructed objects is based on smearing the initial four-momenta of the stable particles according to the tracker and calorimeter resolutions and efficiencies described previously. The calorimeter towers, as output by the simulation, can be used as input to the FastJet [184] libraries for jet clustering. The time necessary for processing a $t\bar{t}$ with DELPHES3 tuned to

describe the CMS detector is about 9 ms [170].

The processing times required by the FASTSIM and ATLAS frozen showers techniques are at least three orders of magnitude larger than the requirement for an ILD fast simulation. DELPHES3 does fullfill this constraint, however, its wide range of simplifications in terms of the detector description and considered material effects are not compatible with the level of realism needed in an ILD fast simulation.

In SGV, the tracker simulation and reconstruction steps are essentially performed simultaneously thanks to the Kalman filter [181] track fitting approach. The simulation of the calorimeters follows the same philosophy: the shower energy is essentially obtained by smearing the particle's true energy with an input detector resolution. Consequently, in contrast to FASTSIM and ATLAS with frozen showers, no additional reconstruction step is required. This offers a significant advantage regarding the necessary computing time.

Furthermore, SGV also provides user-steerable routines that perform track-cluster associations and, subsequently carry out a basic particle identification procedure.

Therefore, in comparison to FASTSIM and ATLAS with frozen showers, SGV can directly output reconstructed particles without requiring further reconstruction tools. The amount of processing time is of the order of $\mathcal{O}(10)$ ms per event [185] which easily fulfils the ILD requirement and is up to three orders of magnitude faster than the performance of the two programs mentioned above.

The advantages presented by SGV concerning the required computing time, the detector description and response as well as their straightforward implementation informed the decision to designate it as the standard fast simulation program for ILD. The following section provides a more detailed overview of its working principles.

7.3 The SGV Fast Simulation

The present summary of the simulation techniques employed in the ILD fast simulation, SGV, follows the discussions in [6] and [178].

7.3.1 Simulation of Particle Interactions and Decays

In the first stage, the program must determine which outgoing particles will interact in the detector. The minimum momentum that electrons and photons must have in order for these interactions to take place can be specified by the user.

Once a photon conversion or electron bremsstrahlung interaction occurs, the initial particle is flagged as decayed which means that the detector "layers" that the particle would traverse after the interaction occured are discarded and *not* taken into account in the measurement error computation. However, if the energy loss suffered by an electron via bremsstrahlung is small enough, the original particle is kept and marked as undecayed. Moreover, the threshold (in terms of the amount of lost transverse momentum via bremsstrahlung) at which a new

electron should be created to replace the "original" one can also be specified by the user. The radiated photon is always saved.

The particle decays that take place in the detector volume are handled in the following manner: typically, the event generator provides information related to the position of the decay vertex and regarding the momenta of the daughter particles. If the decaying particle is neutral, then the information on the decay vertex position can be used directly, since the trajectory of a neutral particle is not affected in the presence of the magnetic field. In the case of a charged decaying particle, the position of the decay vertex and the momenta of the decay products are adjusted to take into account the presence of the magnetic field since the generator has no knowledge about its strength and orientation.

Both for charged as well as for neutral particles, the distance in the detector that they have traversed before decaying is recorded. In case a particle reaches the calorimeter before decaying it is marked as stable [178]. The electrons and photons that reach the calorimeters are discarded from the layers that they would intersect further out, while muons and hadrons continue.

7.3.2 Simulation of the Tracking System

In SGV, the description of the helicoidal track produced by a charged particle is expressed in terms of five parameters, defined with respect to the detector coordinate system and illustrated in figure 7.1:

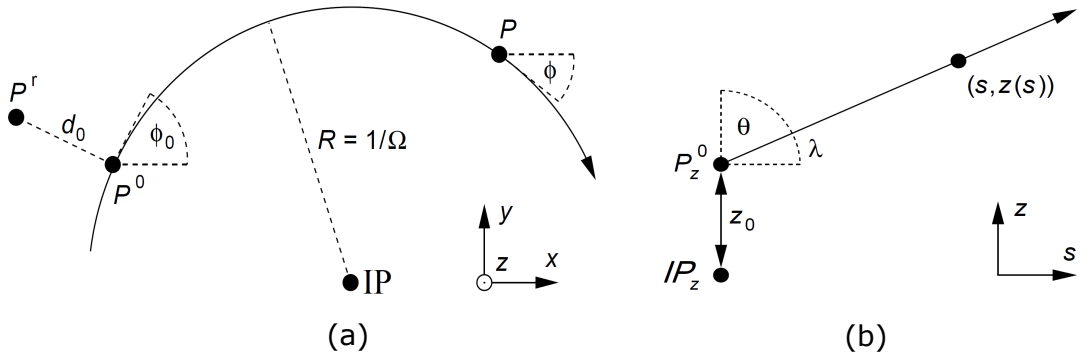


Figure 7.1: Sketch illustrating the parameters used to describe the trajectory of a charged particle in a magnetic field. Panel (a) shows the projection in the x-y plane of a segment of the helicoidal track. Panel (b) shows its projection in the s-z plane, where s represents the corresponding arc length from the x-y plane. Figure adapted from [186].

- d_0, also known as the *impact parameter*, is the distance in the x-y plane between the projection onto that plane of the interaction point (IP) and the perigee (P^0 in figure 7.1).
- z_0, the z-coordinate corresponding to the perigee.
- ϕ_0, the azimuthal angle of the track at the perigee.
- θ_0, the polar angle of the track at the perigee.

→ Ω, defined as $|\Omega|=1/R$, where R is the track radius of curvature in the x-y projection. The sign of Ω is defined by "following" the particle's momentum along the track: it is considered positive if the particle travels counter clock-wise and negative when it moves clock-wise with respect to the IP.

The first step consists of determining the values of the five track parameters at the perigee. This is carried out using the Monte Carlo truth information regarding the momentum of the particle and the position of its starting point in the detector, e.g., the interaction point (IP).

Next, the trajectory along which the particle would travel is extrapolated. The helix is followed throughout the detector and the specific tracker "surfaces" that it intersects are determined and stored in a list. At this point, an approximation is made: all tracks are assumed to originate at the centre of the detector. The tracking continues outwards until the particle's trajectory crosses the starting layer of the outermost calorimeter. This is illustrated in figure 7.2.

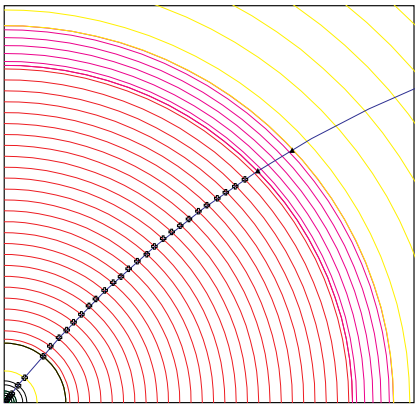


Figure 7.2: A sketch of the tracking simulation procedure. It depicts the projection of a quadrant of the ILD in the x-y plane. The track is followed through the detector and the surfaces that it intersects are recorded. Picture taken from [6].

A sector of the silicon tracking system can be seen in the lower left corner of figure 7.2. It is surrounded by the ILD tracker, here pictured in red. The electromagnetic calorimeter is illustrated by the magenta layers, while the hadronic calorimeter is depicted in yellow. The trajectory of a charged particle, marked in black, is extrapolated through the detector. Its intersections with the tracker layers are illustrated by black markers.

The next step in SGV consists of analytically computing the covariance matrix of the track parameters at the perigee, by using the previously determined list of surfaces that the track intersects. The calculation takes into account both the effects of the extrapolation uncertainties as well as the impact of multiple-scattering and energy loss that can occur when the particle traverses the silicon detector surfaces.

The method applied in SGV for computing the covariance matrix is essentially an adaptation of the well known *Kalman filter* procedure [181]. In the context of track reconstruction, the Kalman filter is a method that can perform both the track *finding* and the track *fitting*. It progresses from one measurement to the next and it improves the knowledge of the particle's trajectory with each added measurement.

In high energy physics, the Kalman filter first materialised in the form of the *Billoir track fitting method* [187], [188]. The principles underlying the two procedures are fundamentally the same and SGV follows the Billoir approach. The following paragraphs provide an overview of how this technique was implemented in SGV.

The covariance matrix (at the perigee) emulates the output and the precision of a track fitting procedure that uses all the measurements performed along the particle's trajectory. For the calculation, the track is followed from the outer-most layer of the tracker inwards. The inward direction is chosen such that the optimal information about the track parameters may be obtained at the closest point possible with respect to the IP [189].

When the track intersects one of the tracker "layers" a new measurement of the track parameters is performed. The addition of the newly determined (track) parameters' values improves the knowledge (i.e., the precision) of the particle's trajectory. The SGV code adds the new values in quadrature to the corresponding elements of the covariance matrix.

When the silicon tracking system is reached, the material effects like multiple-scattering, for example, are taken into account. They can decrease the precision on the track trajectory once again. Therefore, they are added to the *inverse* of the covariance matrix, also known as the *weight matrix*.

After this step, the weight matrix is inverted in order to return to the covariance matrix and the latter is translated along the particle's helicoidal trajectory to the next traversed surface in the list. The procedure is repeated until the point of closest approach (perigee) is reached.

In SGV, the tracker resolution is simulated by using the final covariance matrix for smearing the previously calculated track parameters at the perigee. This is done by means of the *Cholesky decomposition* technique [190]. The Cholesky decomposition is a method for expressing a Hermitian, positive-definite matrix (i.e., in this case, the track parameters' covariance matrix at the perigee) as the product of a lower triangular matrix (L) and its conjugate transpose (L^*).

In order to randomise the smearing process, the lower triangular matrix L is multiplied by a vector $\vec{v}$ which contains uncorrelated random variables [6]. This method ensures that the product $L\vec{v}$ will retain the correlations of the covariance matrix.

As a simplification, any pattern recognition issues that could arise in the real track reconstruction are not taken into account. Problems in the hit pattern recognition could lead to a number of tracks being lost. In SGV, a track is missing only if its covariance matrix cannot be computed. Such a situation arises when the particle's trajectory intersects far too few tracker surfaces (i.e., less than three).

This is essentially the case of too few track measurements which can happen if the charged particle: *(i)* has very low momentum and cannot reach the 3 inner-most layers, *(ii)* travels down the beam-pipe or through an uninstrumented gap in the detector or *(iii)* it decays before hitting the minimum required 3 layers.

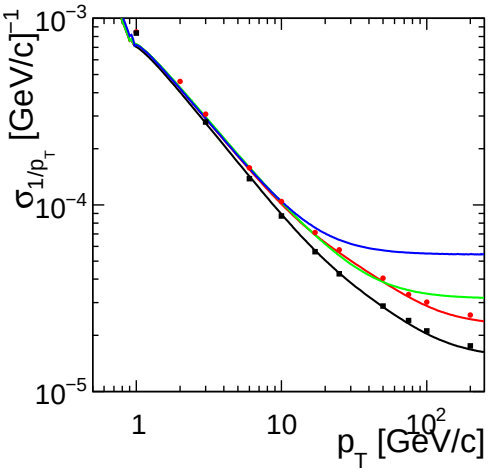


Figure 7.3: Performance of SGV tracker simulation: comparisons of the momentum error $\sigma_{1/p}$ as a function of the particle's transverse momentum between the full simulation (dots) and SGV (lines) for various detector configurations. Figure taken from [6].

The performance of the SGV implementation of the ILD tracking system is illustrated in figure 7.3 as the momentum resolution versus the particle's traverse momentum.

The distribution was obtained for various configurations of ILD, marked by the different colours. The result of the full ILD simulation is presented as the dotted markers, while the SGV performance is depicted by the continuous lines. The very good agreement between the two (simulations) is clearly visible in figure 7.3.

7.3.3 Simulation of the Calorimeters

In order to meet the stringent computing time requirements, i.e., 10 ms for simulating and reconstructing an event, SGV does not simulate the development of a calorimeter shower.

In the first step, the program determines which calorimeters will be reached by the particles. For this purpose, the intersection points between each particle's trajectory and the calorimeters' surfaces are determined. The outcome is a list of reached calorimeters, created for each particle in the event. The list is ordered according to the distance between the point where the particle originates and the calculated intersection point.

Following this stage, SGV must decide for each particle *if* and *how* it will interact in the calorimetric system. This involves first checking whether the particles decay before reaching the calorimeters. If they do not, the electrons and photons deposit their energy in the ECAL, the hadrons are detected as minimum ionising particles

(MIPs) in the electromagnetic calorimeter while the muons are seen as MIPs in both the ECAL and the HCAL [178]. The true Monte Carlo energy of each particle is verified with respect to the user set energy-threshold that determines whether the particle will be detected at all. Finally a detector efficiency is applied which can result in some particles not being seen in the calorimeters. However, even if a particle is not detected in the calorimetric system, SGV still checks whether the particle may be measured later on in a different detector subcomponent.

Once it was decided that a particle reaches and interacts in the calorimeters (either as a MIP or by showering), the detectors' energy response is simulated. In the case of particles that produce a calorimetric shower, the observed energy is determined by sampling a Gaussian distribution that has the true (Monte Carlo) energy of the particle (E_{true}) as its mean. When hadrons start showering before reaching the HCAL, the amount of energy deposited in the respective sub-detector is first subtracted. The measurement uncertainty (σ_E) is given by $\sigma_E = \sqrt{r_1^2/E_{true} + r_2^2}$, where r_1 and r_2 are parameters that must be provided by the user. For MIPs, both the measured energy and the width of the energy distribution must be provided by the user.

In the next step, the shower axis is simulated in SGV. In the case of highly granular detectors, as is the case of ILD, the shower axis is described by four parameters: two for the starting point of the shower and two for its angular orientation, i.e., the polar and the azimuthal angles. The measured values and the associated uncertainties are simulated as described above for the energy response.

One of the most challenging stages in the Particle Flow event reconstruction is the calorimeter shower-track association. SGV provides a routine that can perform this task. The user can input a value for the minimum distance (in cm), between the starting point of the shower and the point where the track extrapolation reaches the calorimeter inner surface, at which it is possible to separate the track from the shower. The program loops over all tracks and showers and decides, depending on each calculated distance, whether they can be associated. As a measure for increasing realism, the routine is designed to accommodate the further implementation of track-shower association erros.

The final stage consists of performing a basic particle identification procedure. Thus, SGV identifies: the showers produced by charged particles in the ECAL as electrons or positrons, the charged particles giving rise to hadronic showers in the HCAL are identified as charged pions, while in the case of neutral particles the showers in the ECAL are associated to photons and those in the HCAL to K_L^0.

Furthermore, SGV also emulates an important component of the Particle Flow reconstruction by storing the tracker measurements for determining the energy of charged particles and the calorimeter measurements for the particles that were not detected in the tracker.

The techniques described above and their straightforward software implementation ensure that SGV fulfils the flexibility, realism and computing speed requirements for a fast simulation of ILD.

The Particle Flow reconstruction is crucial for achieving the physics goals of the ILC. Consequently, SGV must emulate the ILD Particle Flow performance as

determined from full simulation studies. In this regard, the excellent agreement between SGV and the full simulation concerning the tracking performance is very promising.

Evaluating and studying ways of improving the Particle Flow performance of SGV were some of the main tasks of this thesis. The dedicated studies performed and their results will be presented in detail in the next chapter.

CHAPTER 8

Particle Flow Emulation in SGV

SGV, *La Simulation à Grande Vitesse* [6], is the fast simulation program chosen for the International Large Detector (ILD). It fulfils the requirements of a highly granular, Particle Flow based detector to a high degree. Data samples produced with SGV have already been used by a large number of ILC physics studies, e.g., the analysis of a supersymmetry scenario involving light and mass degenerate higgsinos [80].

The SGV simulation of the ILD tracking system agrees very well with the result of the full simulation, as shown in the previous chapter, i.e., in figure 7.3.

Concerning the simulation of the ILD calorimetric system, SGV includes from the start a substantial degree of realism. The program incorporates methods to: *(i)* randomly generate errors on the shower energy, position and shape (i.e., shower axis), *(ii)* combine showers based on their position and axes information and *(iii)* further implement association errors in the track-shower matching procedure.

In order to ensure that the Monte Carlo data produced with SGV can be used for physics and detector optimisation studies, it is essential that the fast simulation emulates the Particle Flow approach and performance of the ILD full reconstruction, specifically, the PandoraPFA [7] reconstruction algorithm, to a high degree.

Evaluating the Particle Flow performance of SGV was one of the central tasks of this thesis. The present chapter first provides a detailed account of the study carried out to determine the default SGV Particle Flow performance. It then discusses the potential association errors that lead to an inevitable degree of confusion in the Particle Flow reconstruction and describes how they have been implemented in SGV. Lastly, the SGV performance after implementing the confusion emulation procedure is presented.

8.1 SGV Default Particle Flow Performance

The Particle Flow performance is typically evaluated in terms of the jet energy resolution, defined and discussed in chapter 4. Following the simplifications described and motivated in sections 4.2 and 6.6.2, the jet energy resolution can be approximated by: $\sigma_{E_j} \approx \sigma_{Det.} \oplus \sigma_{Conf.}$.

The detector energy resolution already implemented in SGV, as presented in section 7.3, accounts only for the $\sigma_{Det.}$ contribution. In the default running mode of SGV, the Particle Flow confusion effects, $\sigma_{Conf.}$, are not taken into account. Thus, the fast simulation perfectly reconstructs the calorimeter showers, without considering any of the unavoidable clustering errors that may occur in the Particle Flow reconstruction. Furthermore, it also performs perfect track-cluster associations. This SGV running mode is known as "SGV *with perfect calorimetry*" (SGV_{PERF}) and it represents a first order approximation of the full reconstruction. Consequently, in view of the physics analyses performed with SGV Monte Carlo data, it is important to investigate how well the PandoraPFA performance is emulated by the SGV_{PERF} output.

For this purpose, a study was carried out in the context of this thesis using Monte Carlo data samples typically denoted as "$Z \rightarrow uds$". As discussed in section 6.6.2, the motivation for using this type of events is that the single jet energy can be directly estimated from the considered ILC centre-of-mass energy: $E_{jet}=\sqrt{s}/2$. This averts introducing inevitable association errors that can occur when employing a jet clustering algorithm.

The SGV_{PERF} output was evaluated in comparison to the results obtained with the two subsequent versions of the ILD full simulation and reconstruction, LoI and DBD, described in chapter 6.

Table 8.1 provides an overview of the data sets used in the performance study. Thus, six different $Z \rightarrow uds$ data sets have been produced with the default (i.e., DBD equivalent) ILD configuration of SGV_{PERF} (version rv86), considering the following ILC centre-of-mass energies: 30, 40, 91, 200, 360 and 500 GeV. Each data sample consisted of 10000 events.

$\sqrt{s}$ [GeV]	LoI	DBD	SGV_{PERF}
30	-	Simulated	Simulated
40	-	Simulated	Simulated
91	Standard	Standard	Simulated
200	Standard	Standard	Simulated
360	Standard	Standard	Simulated
500	Standard	Standard	Simulated

Table 8.1: Summary of the $Z \rightarrow uds$ data sets used in the comparison of the SGV_{PERF} Particle Flow performance. The data samples produced specifically for this study are marked as "Simulated".

Concerning the full simulation, only the four $Z \rightarrow uds$ data sets in the energy range between 91 and 500 GeV were originally produced, for each version.

Based on the data obtained with the LoI full simulation[1], the PandoraPFA performance was evaluated and the results were presented in the ILD *Letter of Intent* document [109]. The DBD data[1] was employed in the assessment of the PandoraPFANew performance and the outcome was described in the "*Detectors*" volume of the ILC *Technical Design Report* [123] (TDR). All these previously obtained performance results are used directly for the comparison with SGV_{PERF}.

Nevertheless, in a significant number of collision events with more than two jets, at least one of them can have an energy lower than 45.5 GeV. Therefore, it is important to evaluate and compare the Particle Flow performance for less energetic jets as well. In the LoI case, the simulation and reconstruction of the two additional lower energy data samples, i.e., $\sqrt{s}$=30 GeV and $\sqrt{s}$=40 GeV, was no longer technically feasible. However, the two data sets have been fully simulated and reconstructed, specifically for this study, using the DBD version.

The SGV_{PERF} performance was analysed in terms of two observables that are highly important for the Particle Flow reconstruction: *(i)* the visible energy, i.e., the sum of the reconstructed particles' energies and *(ii)* the jet energy resolution. The results and their comparison to the full simulation are discussed in the following.

Visible Energy

The distribution of the total visible energy (E_{vis}) reconstructed in $Z \rightarrow uds$ events is shown in figure 8.1 for four different ILC centre-of-mass energies: 91, 200, 360 and 500 GeV. Due to the limited data availability, the SGV_{PERF} result, illustrated by the dashed, coloured line, is compared only to the DBD output.

The E_{vis} distributions have been compared in terms of: their peak position, expressed in terms of Mean_{90}[2] and their width, expressed as rms_{90}[2].

It can be seen from figure 8.1 that, for all four shown SGV_{PERF} distributions, the position of their peak is well compatible with the DBD result. Indeed, this is the case for all six considered data samples. The Mean_{90} values all agree within 1%.

The situation is rather different regarding the widths of the visible energy distributions. It is already noticeable from figure 8.1 that the SGV_{PERF} E_{vis} distributions become narrower with respect to the DBD histograms as the value of the centre-of-mass energy increases.

A comparison of the rms_{90} values reveals that: for the lowest energy sample, i.e., $\sqrt{s}$=30 GeV, the width of the SGV_{PERF} distribution is approximately 7% smaller than its DBD counterpart. This effect decreases up to a factor of three for centre-of-mass energy range between 40 and 200 GeV.

However, in the case of the last two $Z \rightarrow uds$ data sets, i.e., with $\sqrt{s}$=360 GeV and $\sqrt{s}$=500 GeV, the SGV_{PERF} distribution is narrower than the DBD output by 21% and 29%, respectively. The width of the visible energy distributions obtained with the full simulation increases at high jet energies. This is expected since the relevance of the confusion effects increases with the jet energy [7].

The discrepancy is even larger in the comparison with the quoted LoI data:

[1] Marked as "Standard" in table 8.1.

[2] Following the definition presented in section 6.6.2.

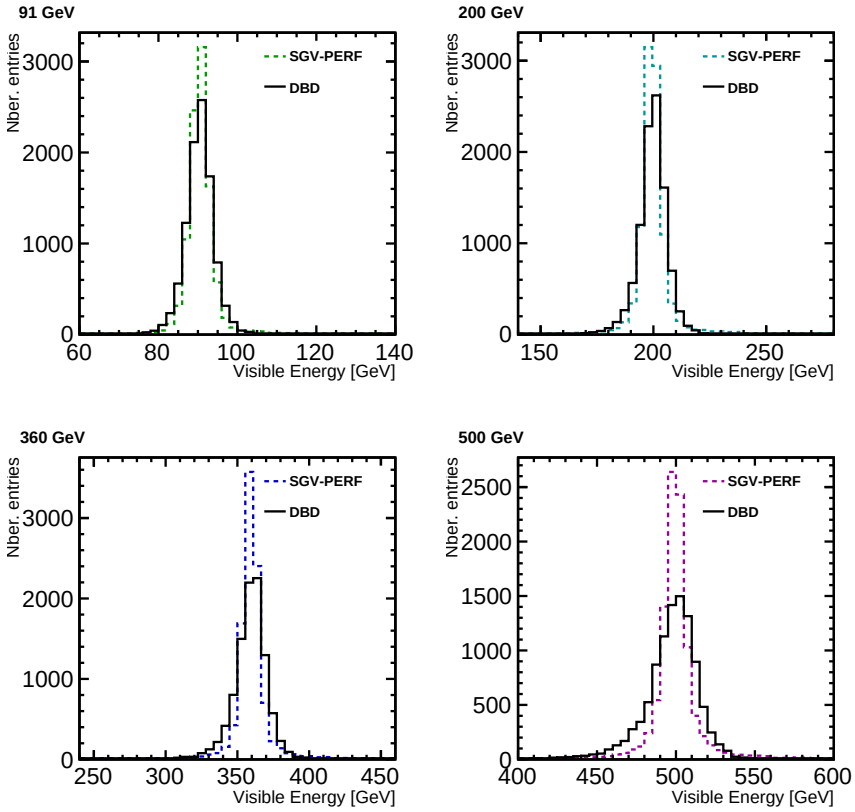


Figure 8.1: Comparison of the $\texttt{SGV}_{\texttt{PERF}}$ performance with respect to the DBD full simulation. The comparison is performed in terms of the visible energy in the reconstructed $Z \to uds$ events for four different ILC centre-of-mass energies.

23% for the $\sqrt{s}$=360 GeV data set and 33% in the case of the $\sqrt{s}$=500 GeV data sample. Thus, from the obtained rms_{90} values it can be concluded that SGV, with the default perfect calorimetry setting, does not fully reproduce the full simulation result.

Jet Energy Resolution

The jet energy resolution is defined, as presented in section 6.6.2, by $\sigma_{\mathrm{E_j}}/\mathrm{E_j} = rms_{90}(\mathrm{E_j})/Mean_{90}(\mathrm{E_j})$. However, what is actually observed is, in fact, the visible energy. In the considered $Z \to uds$ events, the visible energy is equivalent to the di-jet energy: $\mathrm{E_{vis}} \equiv \mathrm{E_{jj}}$. The di-jet energy resolution is described analogously by the expression: $\sigma_{\mathrm{E_{jj}}}/\mathrm{E_{jj}} = rms_{90}(\mathrm{E_{jj}})/Mean_{90}(\mathrm{E_{jj}})$. Thus, using the obtained $rms_{90}(\mathrm{E_{jj}})$ and $\mathrm{Mean}_{90}(\mathrm{E_{jj}})$ values, the di-jet energy resolution can be calculated for the six data samples employed in this analysis.

Furthermore, for the $Z \to uds$ type of events, the link between the jet energy resolution and the di-jet (visible) energy resolution is given by: $\sigma_{\mathrm{E_j}}/\mathrm{E_j} = \sqrt{2} \cdot \sigma_{\mathrm{E_{jj}}}/\mathrm{E_{jj}}$, as shown in section 6.6.2. Consequently, the discrepancies observed between the rms_{90} values of the $\texttt{SGV}_{\texttt{PERF}}$ and the full simulation $\mathrm{E_{vis}}$ distributions will have an impact on the corresponding jet energy resolutions.

In figure 8.2, the jet energy resolution is represented as a function of the sin-

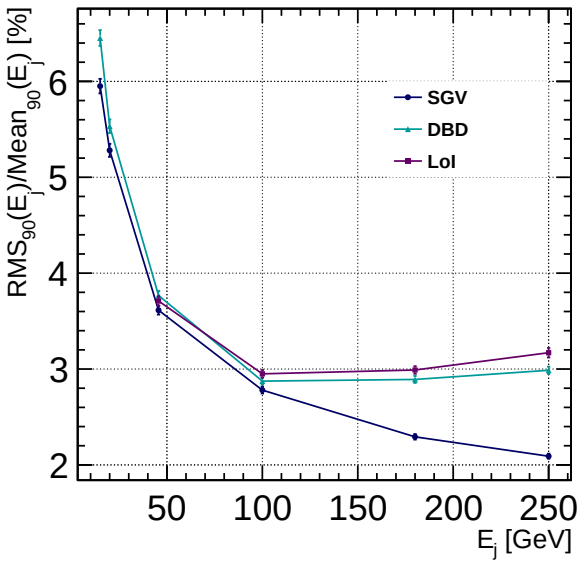


Figure 8.2: Performance of SGV: comparison in terms of the jet energy resolution with respect to the full simulation versions DBD and LoI, described in chapter 6.

gle jet energy.

The small differences that can be observed between the LoI (PandoraPFA) and DBD (PandoraPFANew) results were addressed in section 6.6.2.

It can be seen from figure 8.2 that, for jet energies between 50 and 100 GeV, SGV_{PERF} can well reproduce the jet energy resolution obtained with the DBD version of the full simulation. However there is an $\approx 8\%$ difference for smaller jet energies (<50 GeV). This increases up to 30% for the highest jet energy.

Since the inevitable Particle Flow confusion effects increase with the jet energy it was expected that the discrepancy between SGV_{PERF} and DBD also becomes larger at high jet energies.

However, the $\approx 8\%$ disagreement at low jet energies requires further investigation. There are several potential causes, e.g.: *(i)* the threshold effects that may occur in highly granular calorimeters and that could lead to energy loss if small signals are split over a large number of calorimeter cells or, once again, *(ii)* standard confusion effects.

Lastly, it must be noted that the difference between the SGV_{PERF} and the LoI results is on average up to 5% larger than in comparison to the DBD version.

This implies that, in physics studies, the "perfect calorimetry" mode of operating SGV would provide a too optimistic emulation of the Particle Flow reconstruction. The discrepancy between the SGV_{PERF} and the full simulation results most probably stems from the confusion effects that can occur during the Particle Flow reconstruction and which are not taken into account by default in SGV. Therefore, the level of realism of the ILD fast simulation should be increased in this respect. The most important aspects of the Particle Flow confusion that must be taken into account are described in the following section.

8.2 Association Errors in Particle Flow Reconstruction

In comparison to the LHC, the detector occupancy after a collision event is benign at a lepton collider like the ILC. Nevertheless, even in e^+e^- events, the track and cluster multiplicities can be quite large, of the order of $\mathcal{O}(10^2)$. At high energies, the hadronic jets are even more collimated. Calorimeter showers can overlap, providing the Particle Flow algorithms with a significant challenge when reconstructing clusters, despite ILD's highly granular design. Furthermore, this also increases the probability for track-cluster association errors to occur.

There are two main confusion effects that influence the total reconstructed (visible) energy and, implicitly, the jet energy resolution. They are depicted in figure 8.3.

Double-counting

According to the Particle Flow principle, the energy of a charged particle would be determined exclusively from the tracker measurements. When the particle has enough energy to traverse both the TPC and the calorimetric system, both its track and its calorimeter shower are reconstructed. It is possible that, occasionally, an error occurs during the cluster reconstruction. Instead of including all the energy deposits produced by the charged particle into a single cluster, the algorithm inadvertently produces two (or more) distinct clusters instead. Furthermore, when only one of the two clusters is associated with the corresponding track, the particle's energy is *double counted*.

Specifically, the charged particle contributes to the total reconstructed energy twice: *(i)* via the track measurement and *(ii)* via the split cluster fragment that was not matched to any track and thus is treated as a neutral particle. This situation is illustrated in figure 8.3a.

Energy loss

The opposite effect occurs when the calorimetric shower produced by a neutral particle is wrongly reconstructed as two or more distinct clusters. If a charged particle passes through the detector, at a close distance, a fraction of the split cluster (or even the entire neutral cluster) can be inadvertently reconstructed as part of the shower produced by the charged particle. This case is shown in figure 8.3b.

The energy determination of the charged particle follows the Particle Flow principles, i.e., it makes use of the tracker measurement, while the calorimeter information is discarded. Consequently, the energy contribution from the wrongly split and subsequently merged fraction of the neutral cluster[3] is lost.

Other confusion effects like, e.g., associating a charged cluster to the wrong track or splitting a neutral cluster without associating any of the split fractions

[3] In the following, the calorimeter clusters produced by a neutral particle will be called *neutral clusters*, while the ones created by a charged particle will be termed *charged clusters*.

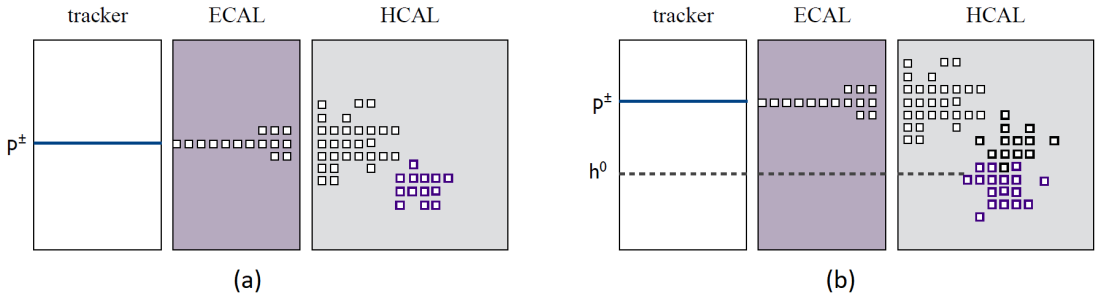


Figure 8.3: Illustration showing the two possible confusion effects that may occur when matching tracks with calorimeter clusters: (a) energy double counting, (b) lost energy. Here, the notation $p^{\pm}$ denotes a charged hadron while h^0 indicates a neutral hadron. Sketch adapted from [108].

to a track, can also occur. However, this type of errors do not impact the total reconstructed energy in the event or the momentum. Therefore, these association errors are not considered in the following discussion.

In order to achieve an SGV output compatible with the full simulation, in terms of total visible energy and, implicitly, jet energy resolution, the *double-counting* and *energy loss* effects must be implemented in the fast simulation. The most effective way to realise this is by studying and parametrising the behaviour of the full reconstruction with respect to the unavoidable Particle Flow confusion. The obtained parametric functions represent a simplified description of the PandoraPFA performance and can be added to SGV. For this purpose, several studies of the PandoraPFA behaviour were carried out. They are described together with the obtained results in the following section.

8.3 Particle Flow Confusion Studies in Full Simulation

The inevitable association errors that can occur in any Particle Flow reconstruction algorithm, as described above, must be emulated in SGV. This endeavour is an ongoing process.

Due to the stringent computing time requirements that a fast simulation must fulfil, SGV does not simulate the development of calorimetric showers and, hence, the confusion errors cannot be implemented at calorimeter shower level.

A much more advantageous solution is to parametrise the likelihood for an association error to occur as a function of particle-level observables, e.g., the particle energy, etc. This can be achieved by performing a detailed study of the full simulation behaviour in terms of several relevant observables. Such a study was carried out for this thesis as an initial proof-of-principle.

8.3.1 The Cluster Merging Probability

The inadvertent merging of calorimeter clusters during the Particle Flow reconstruction stage can have a significant impact on the final visible energy in the event if the merged clusters were produced by particles of different electric charge, i.e., neutral and charged. This confusion error can lead either to the loss of a certain amount of energy in the event or to its double-counting, as illustrated in section 8.2. It is therefore important to investigate the circumstances in which this issue occurs.

It was considered that calorimeter clusters are typically wrongly merged by `PandoraPFA` in events with high particle multiplicities and densities since a very busy environment would be more likely to cause confusion. The degree of isolation of a certain particle and of the cluster it produces in the detector was asumed to affect the likelihood that the latter is accidentally merged with other clusters.

A Monte Carlo study was carried out in the context of this thesis to investigate the relation between the probability that two clusters, produced by two particles of different electric charge, are merged and the degree of isolation exhibited by each of the two particles producing the clusters. The most important aspects and results of the study are presented in this section.

Since confusion errors are more likely to happen in busy events, a Standard Model process characterised by multi-jet final states, i.e., $e^+e^- \rightarrow q\bar{q}q\bar{q}$ ($W^+W^- \rightarrow$ *hadronic jets*) was chosen for the study. A data set of 371000 Monte Carlo events produced with the `LoI` version of the ILD full simulation, considering the ILC running at a centre-of-mass energy of $\sqrt{s}$=500 GeV, was used in the analysis. The reason for using only `LoI` data is that the `DBD` version of the full simulation was still under development at the time when the present study was performed.

In view of a straightforward implementation in `SGV` but also to prevent increasing the required computational time, the confusion effects should be studied and parametrised in terms of observables that are easily accessible in the fast simulation. Consequently, the degree of isolation for any given particle was expressed in terms of the distance, $d_{iso.}$, to the closest neighbouring particle of different charge. From an experimental point of view, there are various approaches to defining the distance between two particles, i.e., their trajectories in the detector. For instance, it is usual for hadron collider experiments to define the distance between two particles in the η-ϕ plane as: $d_{had.} = \sqrt{(\eta_1 - \eta_2)^2 - (\phi_1 - \phi_2)^2}$, where $\eta_{1,2}$ represent the pseudorapidities and $\phi_{1,2}$ the azimuthal angles of the two particles. The distance definition employed in this study is described below.

The Distance Definition

The distance d between any two particles in the event was defined as the three-dimensional straight line distance computed in the ILD coordinate system [191] between two chosen points on the particles' trajectories, as shown in figure 8.4. For simplicity, the location on the particle's trajectory chosen for calculating d was the intersection point, e.g., $\mathrm{P}_i^{int}(x_i,\ y_i,\ z_i)$, where the trajectory reaches the ILD tracker's (TPC) outer cylindrical surface. Another possible choice would have been the point of intersection with the inner surface of the electromagnetic

calorimeter (ECAL).

However, the coordinates for neither of these two potential trajectory locations were stored in the LoI Monte Carlo data. Consequently, this information had to be computed from first principles for this study. The geometrical layout of the ILD ECAL is an octagonal prism which makes it computationally more complex to determine the coordinates of the intersection point with particle's trajectory. In contrast, the ILD tracker is a cylindrical time projection chamber, hence, the calculation of the intersection point is more straightforward. This informed the decision to define d as the distance between the two points where the trajectories of any pair of particles in the event cross the TPC outer surface. For illustration, two such intersection points are marked as $\mathrm{P}_i^{int}(x_i, y_i, z_i)$ and $\mathrm{P}_j^{int}(x_j, y_j, z_j)$ in figure 8.4. The distance is then calculated as: $d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$.

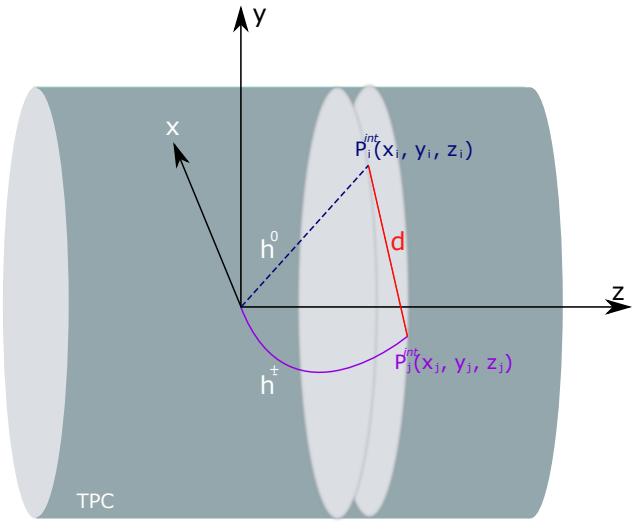


Figure 8.4: Sketch illustrating the definition of the three-dimensional straight line distance between two particles employed in the study of the cluster merging probabilty. All points and distances are defined in the ILD coordinate system.

A procedure was developed for this analysis to determine the coordinates of the intersection point with the TPC outer surface, P_i^{int}, for any particle i. It relies on the Monte Carlo truth information regarding the particle's charged or neutral nature, the location of the vertex where it was created (typically the interaction point) and the magnitude of the three components ($|\vec{p}_x|$, $|\vec{p}_y|$, $|\vec{p}_z|$) of its momentum vector expressed in the ILD coordinate system [191]. Based on this information, the standard five track parameters defined in [192] are determined. The procedure then extrapolates the particle's trajectory until it reaches the TPC outer surface, handling charged and neutral particles separately.

In the case of neutral particles, the procedure simply calculates the intersection of the straight line trajectory with the cylindrical TPC surface.

The trajectories of charged particles are extrapolated as perfect helices, following the calculations described in [193], [192] and [194]. The method differentiates between: *(i)* trajectories that curl inside the TPC volume until they reach the

end-plate and *(ii)* trajectories that intersect the TPC outer surface before completing a full 2π rotation in the x-y plane. The two cases are treated distinctly.

The performance of the trajectory extrapolation procedure is presented in figure 8.5. The shown plots were obtained by drawing the coordinates of the calculated intersection point for each particle in the Monte Carlo data set (371000 $e^+e^- \rightarrow q\bar{q}q\bar{q}$ events) used for the analysis.

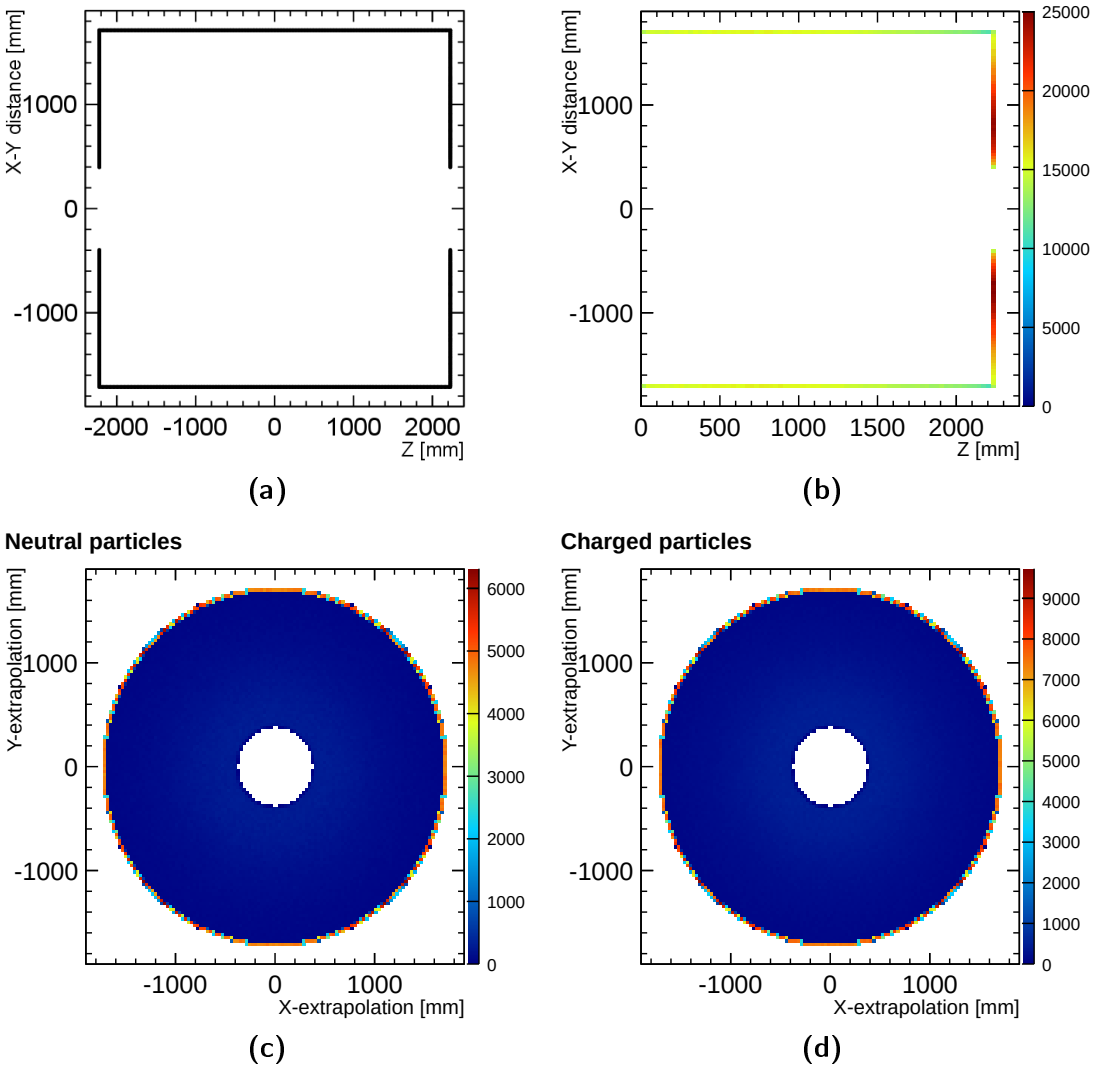


Figure 8.5: Performance of the trajectory extrapolation procedure. The calculated intersection points with the TPC outer surface are shown for the entire data sample in the R-z plane in pannels (a) and (b) and in the x-y projection in pannels (c) (neutral particles) and (d) (charged particles).

The projections in the R-z plane of the calculated intersection points, P_i^{int}, are shown in figure 8.5a, where R represents the radial distance defined as $R_i = \sqrt{x_i^2 + y_i^2}$. The plot contains contributions from both charged and neutral particles. The result essentially shows a transverse section, along the z-axis, of the ILD time projection chamber. The vertical segments correspond to the TPC end-

plates. The dimensions are in very good agreement with the ILD design values quoted, for instance, in figure 5.1. Since the goal of the extrapolation procedure was to compute the intersection points of the particles' trajectories with the outer surface of the ILD tracker, figure 8.5a confirms the accuracy of the computation.

Nevertheless, it can be observed that no intersection points were calculated inside the TPC volume, e.g., for particles interacting in the central cathode plane or in the forward tracking detector (FTD). These detector sub-components were neglected in the extrapolation. The reason for this is that it would be very unlikely for particles interacting in these detector regions to produce relevant calorimeter clusters. Therefore, in the context of the cluster merging probability study, these cases can be safely ignored.

The same R-z projection for both charged and neutral particles is shown in figure 8.5b for one half of the ILD tracker. The colour code indicates the number of particles that have their calculated intersection point in a specific region of the detector. The fact that the study was performed using $W^+W^- \rightarrow hadrons$ events which are typically forward oriented in the detector is illustrated by the red areas (indicating high particle multiplicities) located on the end-plate. This represents another confirmation of the good performance provided by the trajectory extrapolation procedure.

The projections of the calculated intersection points in the x-y plane are shown in figure 8.5c for neutral and in 8.5d for charged particles. The particle multiplicities are illustrated by the colour code. Most particles appear to be associated with an intersection point located at the outermost region of the x-y circular projection. This is due to the fact that the multiplicity values shown in the x-y plane are in fact the result of integrating the numbers of particles having their intersection point in the barrel all along the TPC length in z.

Lastly, it must be noted that the trajectory extrapolation procedure is only a first order approximation since it relies exclusively on geometrical calculations. No energy loss, multiple scattering or magnetic field distortion effects were taken into account. The requirement for a more realistic approach was already met in the subsequent version of the ILD full simulation, `DBD`, by storing the coordinates of several relevant points along a particle's trajectory in the output data.

The Cluster Merging Condition

The `LCIO` event data model (EDM), described in section 6.1, stores not only Monte Carlo truth and reconstruction-related information, but also several links (`LCRelations`) between the two as shown in figure 6.1. The availability and use of these links was crucial for this analysis. The most relevant relations were the ones connecting (in both directions) the information regarding Monte Carlo True particles (`MCPs`) to the associated reconstructed particles (`PFOs`) and their respective clusters formed by `PandoraPFA`.

To illustrate the conditions in which two clusters are considered *merged*, it is useful to refer to the example shown in figure 8.6.

On the Monte Carlo truth level (8.6(a)), two hadrons, a charged ($\texttt{MCP}_\texttt{A}$) and a neutral one ($\texttt{MCP}_\texttt{B}$), interact in the ILD tracker and/or in the calorimetric system.

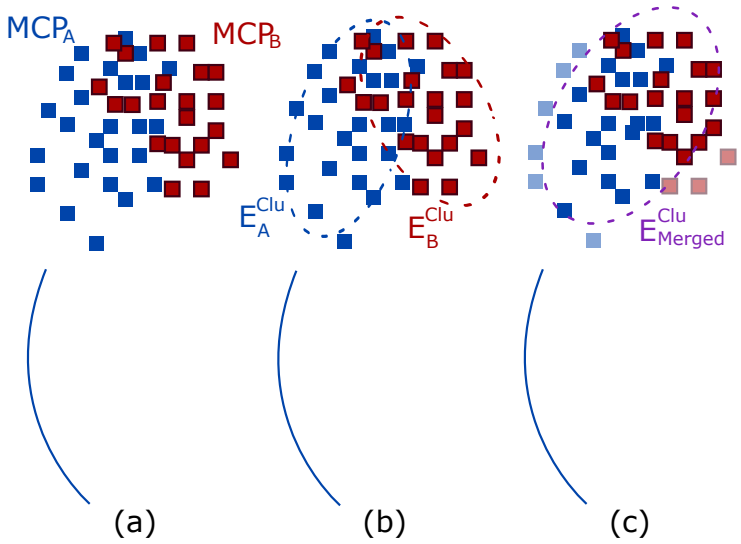


Figure 8.6: Sketch illustrating the accidental merging of two calorimetric clusters in the Particle Flow reconstruction.

Their calorimeter showers overlap. By making use of the chain of `LCRelations` between the simulated calorimeter hits and the formed (reconstructed) clusters, the total amount of *visible* calorimetric energy produced exclusively by the two particles, $\mathrm{E}_A^{Clu.}$ and $\mathrm{E}_B^{Clu.}$, respectively, can be determined.

Due to the inevitable Particle Flow confusion in a busy event, on reconstruction level (8.6(b)), the charged and neutral clusters can be accidentally merged. The result, illustrated in 8.6(c) in this example, is a charged `PFO` with an associated track (coming from $\mathtt{MCP_A}$) and a merged cluster containing contributions from both $\mathtt{MCP_A}$ and $\mathtt{MCP_B}$. The sizes of the individual contributions to the total cluster energy determine the merged status of the two particles.

The fractions of the *original* total calorimetric energies, $\mathrm{E}_A^{Clu.}$ and $\mathrm{E}_B^{Clu.}$, that end up in the merged reconstructed cluster are denoted by: $\mathrm{f}_A^{init.}$ and f_B^{init}. The percentage of the total energy of the *merged cluster* ($\mathrm{E}_{Merged}^{Clu.}$) coming from each of the contributing `MCP`s is denoted as f_A^{merged} and f_B^{merged}, respectively. In this study, two clusters were considered merged if, for both the charged and the neutral hadron the conditions:

$$\begin{cases} 0.2 < \mathrm{f}_i^{init.} < 0.9 \\ 0.2 < \mathrm{f}_i^{merged} < 0.9 \end{cases} \text{(where } i \text{ stands for } A \text{ or } B \text{)} \tag{8.1}$$

were simultaneously fulfilled. This means that, for both Monte Carlo particles, charged and neutral, if a fraction ($\mathrm{f}_i^{init.}$) between 20% and 90% of their individual calorimetric energy ($\mathrm{E}_i^{Clu.}$) contributes to at least 20% and at most 90% of the energy of the merged reconstructed cluster ($\mathrm{E}_{Merged}^{Clu.}$), then the two particles are declared *merged*.

The cases when more than two calorimeter clusters are accidentally merged by `PandoraPFA` were not considered.

The Cluster Merging Probability

The probability, P_{Merged}, for two clusters to be merged was defined as a function of the distance d (following the definition described above) between the two particles that produced the clusters. It is calculated as the ratio between the number of particles (N^{Merged}) having a cluster merged with their closest neighbour of a different charge at a certain distance $d_{iso.}$ and the total number of particles (N^{All}) having their closest neighbour at distance $d_{iso.}$:

$$P_{Merged}(d_{iso.}) = \frac{N^{Merged}(d_{iso.})}{N^{All}(d_{iso.})} \tag{8.2}$$

In this study, the cluster merging probability was determined based on the whole Monte Carlo data sample. Thus, for instance, the denominator in equation 8.2 expresses how many Monte Carlo particles in the entire data sample have their closest neighbour at a given distance $d_{iso.}$. The algorithm developed for determining the cluster merging probability proceeds as follows:

- → First, all simulated Monte Carlo particles (MCPs) must undergo a selection process that: *(i)* rejects the cases when the transverse momentum of the MCP is below 2 GeV and *(ii)* discards the particles that are very forward oriented ($\theta < 8°$). Furthermore, for simplicity, only stable MCPs, i.e., that do not decay in the simulation, were considered. The average particle multiplicity per event is approximately 20 MCPs after applying the cuts.

- → The procedure to extrapolate the particles' trajectories is then applied for every MCP that survives the selection. The individual intersection points with the TPC outer surface are calculated. Based on this information, each MCP is paired to its closest neighbour and the distance, $d_{iso.}$, between them is stored. The pairing is valid only if both particles travel and have their calculated intersection points in the same detector region, i.e., either barrel or end-caps. The more complex cases when, e.g., one trajectory intersects the TPC outer surface in the barrel while the other crosses it in one of the end-caps are discarded.
  By counting how many times a certain value of $d_{iso.}$ appears between two neighbouring MCPs in the whole data sample, the denominator of the probability equation, $N^{All}(d_{iso.})$, is determined.

- → The algorithm to determine the merged status is then run for each pair of closest neighbours. Analogously, by counting how many times the calorimetric clusters produced by two neighbouring MCPs, located at a distance $d_{iso.}$ from each other, are merged, the numerator, N^{Merged}, of the probability function is obtained.

The cluster merging probability was computed separately for particles interacting in the detector's barrel region from the case when they interact in the end-caps. This is motivated not only by the different layout of the two detector regions

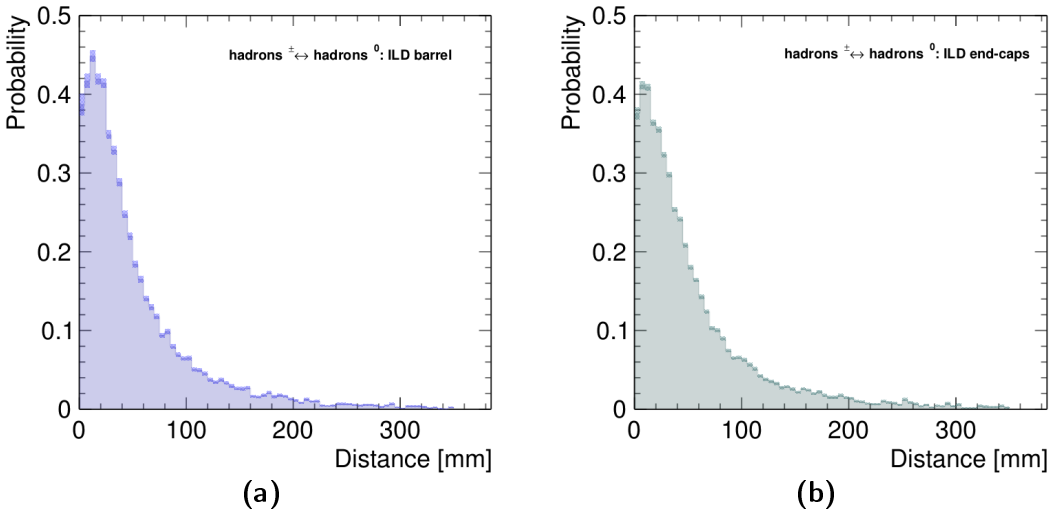


Figure 8.7: The determined cluster merging probability as a function of the distance between the two Monte Carlo particles that produced the two clusters. The size of the binning reflects the size ($5\times5\,\mathrm{mm}^2$) of the ILD ECAL tile. The case when the trajectories of the two neighbouring MCPs intersect the TPC outer surface in the barrel is shown in (a) while the merging probability for MCPs interacting in the end-caps is shown in (b).

but also in view of meeting specific implementation details required by SGV. Both results are shown in figure 8.7.

It can be seen from figure 8.7 that, independent of the detector region, in approximately 40% of the cases when the distance between a charged hadron and its neutral closest neighbour is smaller than 10 cm the two clusters are merged. Furthermore, the merging probability decreases significantly as the distance between the particles increases until it becomes zero for separations larger than 35 cm.

Furthermore, it can also be observed that the highest cluster merging probability is obtained when the two hadrons are separated by a minimum distance of approximately 5-10 mm. This is most likely the effect of the axial 3.5 T magnetic field which encompasses ILD. Even if the charged and neutral hadrons initially travelled through the detector almost collinearly, due to the bending effect of the magnetic field upon the charged particle's trajectory, its calculated intersection point with the TPC outer surface would be "shifted" away from the neutral straight-line trajectory. Furthermore, the charged particle's trajectory continues to bend in the calorimeter as well. The fact that this effect appears more prominent in the barrel region, i.e., where the trajectory bending effect of the B field is larger supports this assumption.

This clear dependency of the cluster merging probability on the degree of isolation of the two particles confirms the initial basic assumption, i.e., that the confusion effects can be parametrised in terms of relevant observables and implemented as an additional function in SGV. However, in this particular case, the study described above was only a proof-of-principle endeavour. Its scope was limited in terms of the number of relevant observables considered and it included

several simplifications, e.g., the distance definition and trajectory extrapolation, etc., that did not meet the requirements of the ILD fast simulation program.

Consequently, a more involved study of the PandoraPFA cluster confusion effects was carried out [6]. Its observations and their subsequent implementation in SGV are summarised in the following section.

8.3.2 The Cluster Splitting Probability

The probability for the Particle Flow reconstruction to wrongly split and merge calorimeter cluster fragments was studied [6] using 8000 events from the same $e^+e^- \to q\bar{q}q\bar{q}$ data sample, fully simulated with the LoI version and reconstructed with PandoraPFA.

The probability for a calorimeter cluster to be accidentally split and merged was investigated as a function of the *true* initial cluster energy and the degree of isolation of the Monte Carlo particle (MCP) that produced it. As in the previous study, the degree of isolation of an MCP was defined in terms of the distance to its closest neighbour of a different charge. The situations when both neighbours are either charged or neutral typically affect neither the total visible energy in the event nor the jet energy resolution. Therefore, such cases were not considered in the analysis.

The distance definition is different from the one employed in the previous study. Thus, for any pair of MCPs, the distance is calculated between their respective shower start positions in the calorimeters. The choice was motivated by the fact that this information is readily available in SGV, making the potential implementation of the confusion parametrisation much easier. Furthermore, the calorimetric showers do not always start at the interface between the tracker (TPC) and the electromagnetic calorimeter.

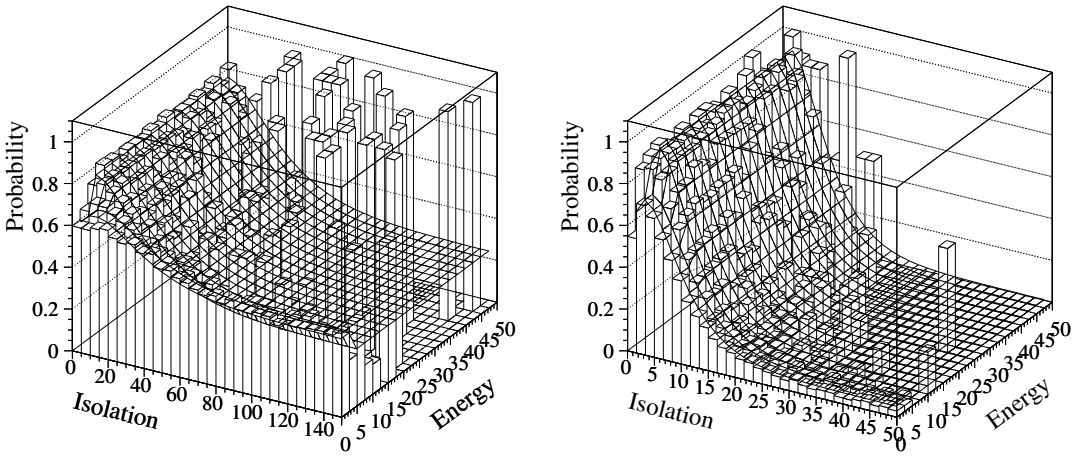
If two MCPs both interact in the barrel region of ILD, the distance between them is expressed in R-ϕ and z coordinates, where R is the radial distance ($R = \sqrt{x^2 + y^2}$) and ϕ is the azmiuthal angle in the ILD coordinate system. At the same time, when the particles interact in the end-cap region, the distance between them is expressed in the x-y plane. The cases when one MCP interacts in the ILD barrel while the other is observed in the end-cap are discarded, as in the previous study.

In order to investigate the relevant cluster reconstruction confusion effects, the same links (LCRelations) between the Monte Carlo truth and the reconstruction-level information were used as in the study described previously.

Two distinct situations were considered in the study: *(i)* the case when several *charged* clusters are wrongly split (and, subsequently, incorrectly merged) which leads to an excess of visible energy in the event and *(ii)* the accidental splitting (and inadvertent merging) of several calorimeter showers produced by *photons* which can lead to a certain amount of visible energy being lost in the event. In contrast to the previous analysis, the clusters produced by neutral hadrons were not investigated.

The obtained results are shown in figure 8.8: the probability that a cluster is

accidentally split is presented as a function of its energy and isolation for charged hadrons in 8.8a and for photons in 8.8b. The figures include contributions from particles interacting both in the ILD barrel and end-caps.



(a) Charged hadrons **(b)** Photons

Figure 8.8: The cluster splitting probability determined as a function of isolation, i.e., the distance to the closest neighbour of different charge measured in cm and cluster energy expressed in GeV. The observed dependency for charged hadrons is shown in 8.8a while the case for photons is presented in 8.8b. Both figures taken from [6].

It can be observed that, in both cases, the cluster splitting probability depends much more on the distance (isolation) between the charged hadron and its closest neighbouring neutral particle than on the energy of the cluster. Furthermore, this dependency is significantly more prominent in the case of photons.

It must be noted that the splitting probability and, implicitly, the likelihood for confusion to occur is high, i.e., above 60%, even for particles with energies below 5-10 GeV, when the degree of isolation is reduced. Since in jets with low energy the individual particle energies cannot be too high either, this appears to be in accordance with the jet energy resolution results presented in section 8.1.

Each of the two distributions was fitted with a combination of linear and exponential functions, illustrated by the mesh in figure 8.8. The fitting functions and their determined parameters were then implemented in `SGV`.

The dependencies presented in figure 8.8 express how likely it is for a charged or neutral (photon) cluster, respectively, to be accidentally split by `PandoraPFA` as a function of the cluster's visible energy and the degree of isolation of the Monte Carlo particle that produced it. The question of how much of the original cluster is actually wrongly reconstructed was also addressed, as described in the following.

The study revealed that, for cases when the *entire* cluster is wrongly attributed, the probability for such an error to occur depends only on the `MCP`'s energy [6].

When only a fraction of the cluster is accidentally separated, two issues must be determined: *(i)* the amount of energy of the split fraction and *(ii)* the probability

that the error occurs for fragments of that particular energy. Figure 8.9 shows the cluster energy fraction *correctly* attributed ($f_{Corr.}$), for both charged hadrons and photons, as a function of the particles' isolation and cluster energy. Evidently, the cluster fraction *wrongly* attributed is given by: $f_{Wrong} = 1 - f_{Corr.}$.

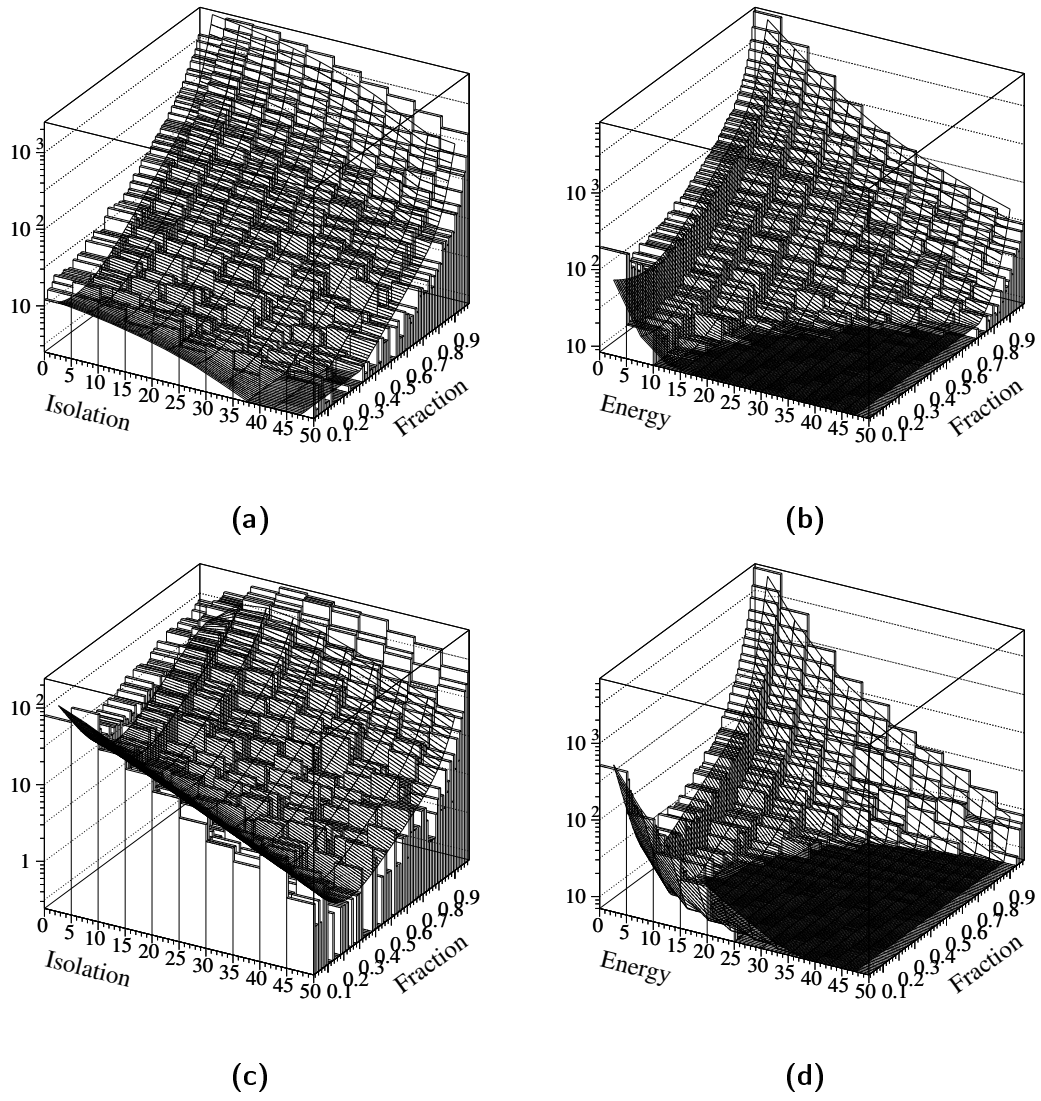


Figure 8.9: The dependency of the fraction from the initial cluster that is *correctly* reconstructed on the degree of isolation (left column) and energy (right column) of the original Monte Carlo particle that produced it. The third (vertical) axis represents the number of entries in the histogram. The upper row shows the situation for charged hadrons while the lower one illustrates the situation for photons. The isolation distance is given in cm and the energy is expressed in GeV. All figures taken from [6].

It can be observed from figure 8.9 that the magnitude (in terms of energy) of the split cluster fragment depends both on the energy and isolation of the `MCP` that created the original cluster. Furthermore, the dependency is more pronounced in the case of photons.

A combination of linear and exponential functions was used to fit each one of

the four distributions shown in figure 8.9. The fit results are represented by the gray hash-lined surfaces. Based on these results, a probability density function was constructed. It expresses the likelihood for a cluster produced by a Monte Carlo particle having its closest neighbour at distance d to have a fraction f_{Wrong} of its energy wrongly attributed. It was also found [6] that the probability density function can be formulated, more conveniently for the fast simulation, in terms of the *average* fraction, where the average is determined over bins of energy and isolation. Thus, the probability density function maintains its dependency on the MCPs' energy and degree of isolation but only by means of the computed average fraction.

Implementation of the Particle Flow Association Errors in SGV

The first attempt to implement the unavoidable Particle Flow confusion effects in SGV is described in the following.

All the fitted functions shown above, together with their respective parameter values, were implemented in SGV via a new routine called ZACCON. In SGV, the simplification that hadrons do not interact in the electromagnetic calorimeter and that electromagnetic showers do not leak into the hadronic calorimeter is applied. Thus, the program first determines, for each Monte Carlo true particle in the event, the distance to the closest neighbour of the opposite charge, i.e., pairs of charged-neutral particles, *only* in the same calorimeter.

Once the closest neighbour was found, the probability functions that determine whether the whole cluster or only a fraction of it will be wrongly split are randomly sampled in order to decide which error must be simulated. If the outcome was that only a fragment of the cluster is wrongly attributed, the program then calculates the size (in terms of energy) of the split fragment.

In the case of a neutral cluster, the simulation attempts to incorporate its split fragment with the cluster of the previously determined closest charged neighbour. In order for the attempt to be successful, the energy sum of the charged cluster plus the contribution from the neutral fragment must be compatible with the track momentum. If that is not the case, the program makes up to ten more attemps with different sizes of the neutral cluster fragment. If the incompatibility remains unchanged, the confusion error is *not* simulated and only the detector resolution is applied to the original energy of the particles.

When a charged cluster is divided according to the computed energy fraction, the split fragment is usually "placed" as a new neutral cluster in the middle of the distance between the charged MCP and its closest neutral neighbour. The simulation then checks if the remaining charged cluster energy is compatible with the track momentum. In case of incompatibility, the program makes up to two more attempts with different charged cluster fragments. If no compatible match could be found, the error is not simulated.

Once the ZACCON routine fulfils these tasks the rest of the simulation steps are taken over by the usual SGV functions. The intrinsic detector resolution is also taken into account, together with the confusion effects emulated by the new routine.

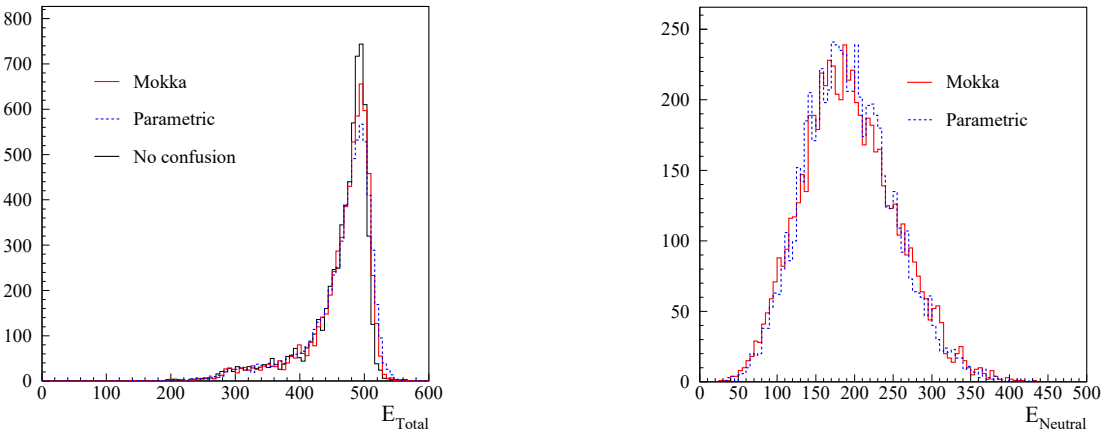
It must be noted that the use of the ZACCON routine is entirely optional. The user can decide which type of calorimeter simulation should be performed by SGV: either considering perfect clustering ($\texttt{SGV}_{\texttt{PERF}}$) or using the Particle Flow confusion emulation as described above. For the rest of this thesis, the latter case will be denoted as $\texttt{SGV}_{\texttt{PFL}}$.

First Performance Evaluation

The Particle Flow confusion effects that can occur during the PandoraPFA cluster reconstruction stage have been studied, parametrised and implemented in SGV, as presented above. In order to validate the realised confusion emulation, it is crucial that the performance achieved with $\texttt{SGV}_{\texttt{PFL}}$ is analysed both with respect to the initial $\texttt{SGV}_{\texttt{PERF}}$ outcome as well as in comparison with the full simulation results.

The first performance evaluation of $\texttt{SGV}_{\texttt{PFL}}$ was carried out in the context of the studies presented in [6]. For this purpose, a sample of Standard Model $WW \to qqqq$ Monte Carlo events, produced at the ILC centre-of-mass energy of $\sqrt{s} = 500\,\text{GeV}$, was considered. For comparison, the same events were processed with: the LoI version of the ILD full simulation, the fast simulation without confusion, i.e., $\texttt{SGV}_{\texttt{PERF}}$ and $\texttt{SGV}_{\texttt{PFL}}$.

The performance was evaluated in terms of the total visible energy in an event and the total reconstructed energy coming from the neutral PFOs in the event. The results are shown in figure 8.10.



(a) Total visible energy

(b) Total energy of reconstructed neutral particles

Figure 8.10: Results of the first evaluation of the confusion implementation in $\texttt{SGV}_{\texttt{PFL}}$. The total visible energy (left) and the total neutral PFO energy (right) distributions were obtained from Standard Model $WW \to qqqq$ events. Figures taken from [6].

In both cases, the distributions obtained with $\texttt{SGV}_{\texttt{PFL}}$ were found to be well compatible with the full simulation. However, comparing the widths of the total visible energy distributions (figure 8.10a) it was observed that the $\texttt{SGV}_{\texttt{PFL}}$ distribution is wider [6]. This implies that, in the context of a physics study, $\texttt{SGV}_{\texttt{PFL}}$ would provide slightly more pessimistic results.

Nevertheless, the performance achieved with $\texttt{SGV}_{\texttt{PFL}}$ in the first evaluation study

was very promising. This prompted a more detailed investigation to be carried out in the context of this thesis.

8.4 Particle Flow Performance of SGV with the Confusion Emulation

The performance of the ILD fast simulation, SGV, run with the Particle Flow confusion emulation option ($\texttt{SGV}_{\texttt{PFL}}$) was evaluated again in terms of the total visible energy summed per event and the obtained jet energy resolution. For this purpose, six $Z \rightarrow uds$ data samples, considering the same ILC centre-of-mass energy values as in table 8.1, were simulated with $\texttt{SGV}_{\texttt{PFL}}$ (version rv86). Each data set consisted of 10000 events.

The obtained results were compared not only with the $\texttt{SGV}_{\texttt{PERF}}$ outcome presented earlier in this section, but also with the LoI and DBD performance discussed in [109] and [123], respectively.

Visible Energy

The visible energy distributions obtained from the $Z \rightarrow uds$ data simulated with $\texttt{SGV}_{\texttt{PFL}}$ are illustrated by the continuous coloured lines in figure 8.11. For convenience, only the four samples with the highest ILC centre-of-mass energies are shown. The DBD and $\texttt{SGV}_{\texttt{PERF}}$ results are also presented for comparison.

It can be seen that, for all four data sets, the $\texttt{SGV}_{\texttt{PFL}}$ distribution is shifted to higher values in comparison with the other two. The same effect was observed at lower centre-of-mass energies as well. Quantitatively, the comparison in terms of the mean (expressed as Mean_{90}) between the $\texttt{SGV}_{\texttt{PFL}}$ and DBD visible energy distributions reveals that, for low centre-of-mass energies ($\sqrt{s}$ =30 to 40 GeV), the mean of the $\texttt{SGV}_{\texttt{PFL}}$ distribution is shifted to higher values by 1.8 GeV, i.e., a 6% effect. The shift towards higher energies decreases gradually to 2% for the other data samples. A similar discrepancy was observed in the comparison with the $\texttt{SGV}_{\texttt{PERF}}$ results. No data concerning the Mean_{90} of the distributions obtained with the LoI version was available for this comparison.

Comparing the widths (expressed as rms_{90}) of the $\texttt{SGV}_{\texttt{PFL}}$ and DBD distributions shows that the former are 30%-40% (400 to 600 MeV) wider in the case of low centre-of-mass energies. The largest difference was found in the $\sqrt{s}$ = 200 GeV sample with the $\texttt{SGV}_{\texttt{PFL}}$ visible energy distribution being 79% (3.2 GeV) wider than the DBD one. The discrepancy decreases to 37% in the case of the data set with the highest centre-of-mass energy.

The same tendency was observed in the comparison with the LoI results: the $\texttt{SGV}_{\texttt{PFL}}$ was approximately 77% broader in the case of the $Z \rightarrow uds$ sample simulated at $\sqrt{s}$ = 200 GeV centre-of-mass energy and only 30% wider for $\sqrt{s}$ = 500 GeV.

The effect is even more prominent in comparison to the $\texttt{SGV}_{\texttt{PERF}}$ distributions. Thus, for the lower centre-of-mass energy samples, the $\texttt{SGV}_{\texttt{PFL}}$ histograms are up to 46% (≈650 MeV) wider. The largest discrepancy with respect to $\texttt{SGV}_{\texttt{PERF}}$ was

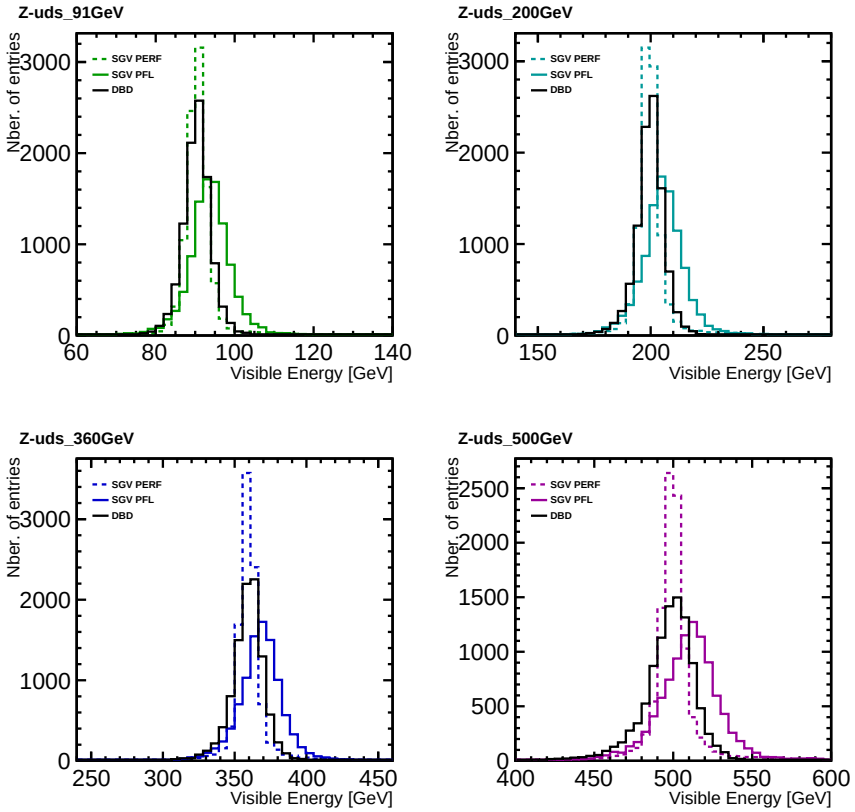


Figure 8.11: Comparison of the visible energy distributions between the DBD full simulation (continuous black line), $\texttt{SGV}_{\texttt{PERF}}$ (dashed coloured line) and $\texttt{SGV}_{\texttt{PFL}}$ results.

observed in the $\sqrt{s} = 500\,\text{GeV}$ sample: the $\texttt{SGV}_{\texttt{PFL}}$ visible energy distribution was found to be 96% (7.1 GeV) wider.

The increase of realism in SGV by means of emulating the Particle Flow confusion effects was indeed expected to lead to wider visible energy distributions. Since the parametric functions were tuned to the performance of the full simulation, the rms_{90} widths of the $\texttt{SGV}_{\texttt{PFL}}$ visible energy distributions were expected to be of the same magnitude as the ones obtained from the DBD and LoI data. However, the results described above show a significant discrepancy between the fast and the full simulation.

The initial performance evaluation of $\texttt{SGV}_{\texttt{PFL}}$ (figure 8.10) also observed that the visible energy distribution is wider than in the full simulation case. However, no shift towards higher energy values was observed. This discrepancy between the two sets of results requires further investigation.

The shift of the $\texttt{SGV}_{\texttt{PFL}}$ visible energy distributions towards higher energy values formed the subject of a dedicated study that will be discussed later in this section. The impact of the confusion implementation on the jet energy resolution is addressed in the following.

Jet Energy Resolution

The jet energy resolution obtained with $\texttt{SGV}_{\texttt{PFL}}$ is presented as a function of the

single jet energy in figure 8.12. The results obtained with $\texttt{SGV}_{\texttt{PERF}}$ as well as with the two versions of the full simulation are also shown for comparison.

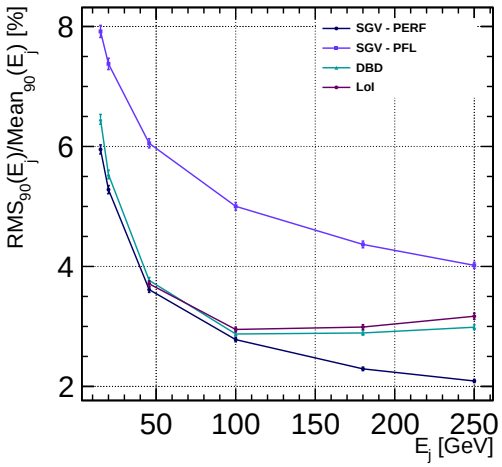


Figure 8.12: Comparison between the $\texttt{SGV}_{\texttt{PFL}}$ Particle Flow performance with respect to the full simulation and $\texttt{SGV}_{\texttt{PERF}}$ results in terms of the jet energy resolution expressed as a function of the jet energy.

In the case of the $Z \rightarrow uds$ events used for this evaluation, the visible energy (E_{vis}) is in fact equivalent to: $E_{vis} = 2 \cdot E_j$, where E_j represents the single jet energy. Thus, it can be observed from figure 8.12 that, as expected, the broader $\texttt{SGV}_{\texttt{PFL}}$ visible energy distributions discussed above are equivalent to an overall worse jet energy resolution.

At low jet energies ($E_j < 100\,\text{GeV}$), the discrepancy between the SGV performance with and without the confusion emulation is of approximately 37%. Furthermore, in comparison to the DBD result, the jet energy resolution obtained with $\texttt{SGV}_{\texttt{PFL}}$ is about 28% worse.

The same behaviour was observed for jet energies between 100 and 200 GeV with $\texttt{SGV}_{\texttt{PFL}}$ providing a worse resolution by approximately 73% in comparison to $\texttt{SGV}_{\texttt{PERF}}$. However, the largest difference between the two, i.e., of the order of 91%, was found for the highest value of the jet energy. The comparison with the full simulation performance for the same value of the jet energy, i.e., E_j=250 GeV, reveals that the $\texttt{SGV}_{\texttt{PFL}}$ jet energy is worse by 43% than the DBD and by 36% than the LoI results.

The Particle Flow confusion parametrisations implemented in SGV clearly affect its performance in terms of both the visible energy observed in an event as well as the achieved jet energy resolution. While most fast simulation programs typically provide too optimistic results, $\texttt{SGV}_{\texttt{PFL}}$ errs on the side of caution by providing rather pessimistic estimates, as shown above. The shift in the visible energy distributions towards higher energy values was investigated in the context of this thesis in a study summarised in the following section.

8.5 The Neutral Visible Energy Correction

The effects of the Particle Flow confusion emulation implemented in SGV manifest in a two-fold manner. The broadening of the reconstructed visible energy distribution and, implicitly, the poorer jet energy resolution can be interpreted and acknowledged as rather pessimistic estimates. However, the inadvertent overall raising of the visible energy (as shown in figure 8.11) leads to a systematic error that must be understood and addressed.

Study of the Visible Energy Shift

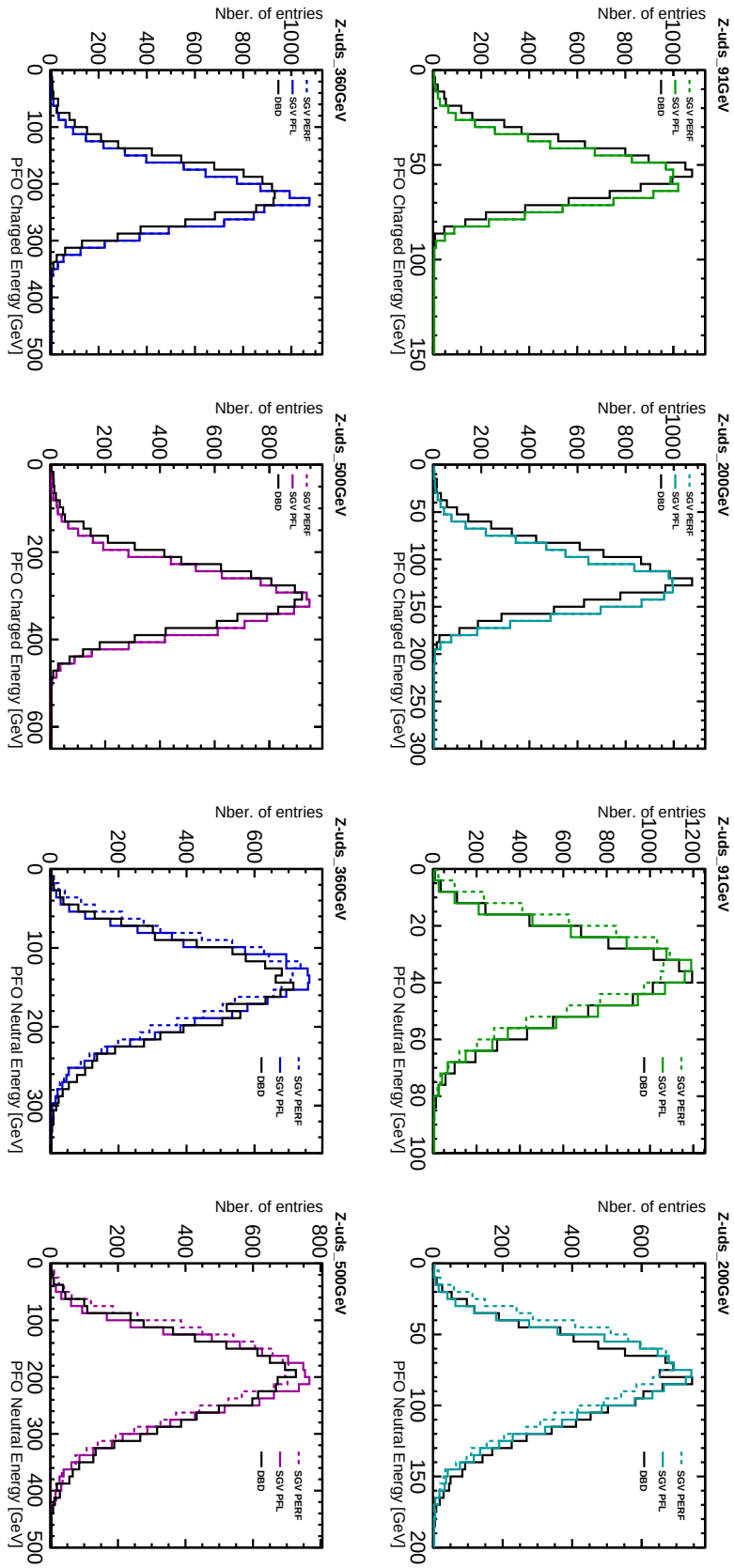
In view of this, three sets of $Z \to uds$ data samples, i.e., simulated with $\texttt{SGV}_{\texttt{PERF}}$, $\texttt{SGV}_{\texttt{PFL}}$ and DBD, considering the six ILC centre-of-mass energies indicated in table 8.1, were employed for a first principles study. For each $\sqrt{s}$ value, 10000 events were produced with each simulation. The same generated events were used in the production of both the $\texttt{SGV}_{\texttt{PFL}}$ and the $\texttt{SGV}_{\texttt{PERF}}$ data samples. No LoI data of this type was available and it was not technically feasible to produce $Z \to uds$ Monte Carlo samples with the LoI full simulation in the context of this thesis. Therefore, the fast simulation performance could only be analysed in comparison with the DBD data.

The study compared several direct observables like, e.g., event shape variables (thrust, sphericity) and the cluster and track multiplicities per event, etc. The most relevant discrepancies were found in the amount of reconstructed *charged* energy, i.e., the total amount of visible energy in the event coming exclusively from the charged reconstructed particles (charged PFOs) and in the amount reconstructed *neutral* energy, defined analogously (i.e., exclusively from neutral PFOs).

The distributions of these two observables, obtained from the data samples mentioned above, are illustrated in figure 8.13 for a selection of four centre-of-mass energies. The following observations were made:

➛ In case of the **charged visible energy** (8.13a), the Particle Flow confusion emulation has no visible effect, i.e., the $\texttt{SGV}_{\texttt{PERF}}$ (dashed coloured line) and $\texttt{SGV}_{\texttt{PFL}}$ (continuous coloured line) distributions overlap. This outcome is understood since, under the Particle Flow paradigm, the energy of the charged reconstructed particles is measured (almost) exclusively from the tracker information which, in the case of $\texttt{SGV}_{\texttt{PFL}}$, is not affected by the confusion parametrisations. Specifically, neither the accidental splitting of a charged cluster nor its inadvertent merging with a neutral cluster would have a major impact on the measurement of the charged particle's energy. If the associated track fulfils the criteria for being reconstructed as a particle on its own, the energy information coming from the related merged or split cluster would be discarded in any case.

It can be observed, though, that both the $\texttt{SGV}_{\texttt{PERF}}$ and the $\texttt{SGV}_{\texttt{PFL}}$ distributions are shifted towards higher energy values in comparison with the DBD (continuous black line). However, the performance of the $\texttt{SGV}_{\texttt{PERF}}$ tracker simulation, illustrated in figure 7.3 in terms of the achieved precision on the transverse momentum, shows a good compatibility with the full simulation. Therefore, the observed excess of



(a) *Charged* reconstructed energy

(b) *Neutral* reconstructed energy

Figure 8.13: The two observables that show the most relevant discrepancies when comparing the behaviours of the three simulations, SGV$_{\text{PERF}}$, SGV$_{\text{PFL}}$ and DBD: the total reconstructed charged visible energy is shown in 8.13a while the case of the neutral visible energy is presented in 8.13a.

charged visible energy is most likely caused either by a too high track reconstruction efficiency and/or by a too large percentage of tracks allowed to form reconstructed particles. A better tuning of the track selection criteria, such that it reflects the behaviour of the `PandoraPFANew` quality checks which determine whether a track can form a reconstructed particle on its own, could probably correct this effect.

➛ Concerning the **neutral visible energy** (8.13b), it was observed that the $\texttt{SGV}_{\texttt{PERF}}$ (dashed coloured line) distribution is shifted to *lower* energy values in comparison with the full simulation (continuous black line). This effect clearly compensates the previously observed excess of *charged* visible energy such that, for $\texttt{SGV}_{\texttt{PERF}}$, the mean of the total visible energy, i.e., charged plus neutral, matches the `DBD` counterpart within 1% (see 8.1). This observation may again be a consequence of the too high track reconstruction (or acceptance) efficiency combined with a too large track-cluster matching rate. Since the cluster energy information is typically discarded once matched with a viable track, an excess of neutral visible energy could only occur when: *(i)* charged or merged clusters cannot be associated with a track, thus, being classified as neutral particles or when *(ii)* charged clusters are inadvertently split and the wrongly reconstructed fragments are considered neutral particles. In $\texttt{SGV}_{\texttt{PERF}}$, the first situation, while implemented [178] does not emulate the behaviour of the full simulation, as can be seen from figure 8.13b.

The second confusion error is implemented only in $\texttt{SGV}_{\texttt{PFL}}$. Indeed, it can be seen that the neutral visible energy distribution obtained with $\texttt{SGV}_{\texttt{PFL}}$ is closer to reproducing the `DBD` behaviour[4]. However, due to the excess of charged energy already present in the event, the extra neutral energy produced in $\texttt{SGV}_{\texttt{PFL}}$ by the charged cluster splitting emulation leads to the *total* visible energy being shifted towards higher values (section 8.4). This is obviously an unwanted systematic error.

In order to counteract its effects, an energy correction procedure was developed in the context of this thesis, as discussed below.

Neutral Visible Energy Correction

The implementation of the unavoidable Particle Flow confusion effects in `SGV` is an ongoing project. Nevertheless, in view of the physics studies to be carried out using Monte Carlo data produced with $\texttt{SGV}_{\texttt{PFL}}$, the present state of the confusion emulation is not satisfactory. While the achieved jet energy resolution can be regarded as a pessimistic approximation, the observed surplus in the visible energy distributions (figure 8.11) must be corrected with a straightforward solution, easily applicable on the analysis level.

The final goal is that, independent of the considered centre-of-mass energy, the $\texttt{SGV}_{\texttt{PFL}}$ distribution is corrected such that its mean value is compatible with the orginal $\texttt{SGV}_{\texttt{PERF}}$ and `DBD` central values. Considering the results presented in figure

[4]This observation appears to be in agreement with the result (figure 8.10b) obtained in the initial $\texttt{SGV}_{\texttt{PFL}}$ evaluation [6]. Consequently, the discrepancy between the two sets of results concerning the visible energy distribution forms the focus of future studies

8.13, this could be achieved by applying a scaling factor, k, either to the charged or to the neutral (or potentially to both) fractions of the total reconstructed energy in the $\texttt{SGV}_{\texttt{PFL}}$ data.

Adjusting the charged component of the visible energy would not be advantageous. The SGV tracker simulation reproduces very well the full simulation performance, as illustrated by figure 7.3. Furthermore, it is not inadvertently affected by the implemented confusion emulation, as can be seen from the overlapping $\texttt{SGV}_{\texttt{PFL}}$ and $\texttt{SGV}_{\texttt{PERF}}$ distributions in figure 8.13a. Therefore, there is no error to be corrected in the SGV simulation of the ILD tracker and, hence, no need to apply a scaling factor to the charged fraction of the visible energy[5].

Nevertheless, the surplus of total visible energy observed in the $\texttt{SGV}_{\texttt{PFL}}$ distributions could be amended by downscaling *only* their respective neutral fractions. For this purpose, the neutral visible energy histograms obtained from data simulated with the confusion emulation (figure 8.13b) must be "shifted" such that their own central values are compatible with their $\texttt{SGV}_{\texttt{PERF}}$ counterparts.

Thus, for every event simulated with $\texttt{SGV}_{\texttt{PFL}}$, the obtained *neutral* visible energy ($\mathrm{E}_0^{vis.}$) must be scaled with a correction factor, k. The specific value of k clearly depends on the total amount of neutral energy reconstructed in the event: $k = k(\mathrm{E}_0^{vis.})$.

The neutral visible energy (hereinafter also referred to as "neutral energy") is determined by summing the energies of all individual *neutral* reconstructed particles, i.e., $\mathrm{E}_0^{vis.} = \sum_i \mathrm{E}_0^{PFO_i}$. Therefore, scaling the neutral energy (the sum) with the factor k is equivalent to scaling the energy of each neutral PFO with the same amount: $k \cdot \mathrm{E}_0^{vis.} = k \cdot \sum_i \mathrm{E}_0^{PFO_i}$. The potential additional dependency of the scaling factor on the individual PFO energy was neglected in this study.

When applying the energy correction on the reconstructed particle level, the mass of the reconstructed particles, i.e., the outcome of the most probable mass hypothesis (section 7.3.3), cannot be modified. Consequently, only the momentum of the particles can be adjusted: $k \cdot \vec{p} = (k \cdot \vec{p}_x, k \cdot \vec{p}_y, k \cdot \vec{p}_z)$.

Since neutral particles are measured only in the calorimetric system, adjusting their momenta can be performed without any implications concerning the performance of the SGV tracker simulation. Scaling the neutral PFO energies with a factor k is analogous to correcting the effects of a systematic bias that affects the calorimetric measurements. Under the Particle Flow paradigm it is possible to apply this energy correction only to the neutral particles since the energies of the charged PFOs are meant to be determined entirely from the tracker information.

Before adjusting the observed energy of each individual neutral particle, the mathematical formulation of the scaling function, $k(\mathrm{E}_0^{vis.})$, must be determined. This can be achieved by comparing the two mean values of the $\texttt{SGV}_{\texttt{PFL}}$ ($\mathrm{Mean}_{\mathrm{PFL}}$)

[5] The excess of charged visible energy observed in figure 8.13a is exhibited identically by *both* variants of the fast simulation. As discussed earlier, it is probably caused by using too optimistic estimates for the track reconstruction efficiency and/or for the percentage of tracks classified as suitable for being reconstructed as charged particles. The overestimation does not degrade the achieved momentum precision of the SGV tracker simulation. Moreover, correcting its effect would require a more involved adjustment of the relevant SGV routines and, therefore, exceeds the scope of this study.

and $\texttt{SGV}_{\texttt{PERF}}$ ($\text{Mean}_{\text{PERF}}$) neutral energy distributions (figure 8.13b), both obtained for the same ILC centre-of-mass energy. Since the goal is to scale the former such that it becomes equal to the latter, i.e., $\text{Mean}_{\text{PERF}} = k \cdot \text{Mean}_{\text{PFL}}$, the scaling factor for that particular amount of reconstructed energy is clearly: $k(\text{E}_{\text{vis}} \equiv \text{Mean}_{\text{PFL}}) = \text{Mean}_{\text{PERF}}/\text{Mean}_{\text{PFL}}$. The energy scaling function can be obtained by calculating this ratio for a number of $\texttt{SGV}_{\texttt{PFL}} \leftrightarrow \texttt{SGV}_{\texttt{PERF}}$ pairs of neutral visible energy distributions.

In total, fourteen $Z \rightarrow uds$ data sets, covering an ILC centre-of-mass energy range from 30 to 500 GeV, have been simulated independently with each of the two variants of the fast simulation. For each data sample, the neutral energy distribution was obtained and fitted with a Gaussian function. The mean values, i.e., $\text{Mean}_{\text{PERF}}$ and Mean_{PFL}, respectively, were extracted from the fits. It was observed that, in all cases, the fitted mean values correspond to almost half of the total visible energy. Their ratio is presented in figure 8.14 as a function of the corresponding Mean_{PFL} values. The fitting errors on $\text{Mean}_{\text{PERF}}$ and Mean_{PFL} were considered in the error propagation carried out to determine the total uncertainty on their ratio, indicated by the error bars in figure 8.14.

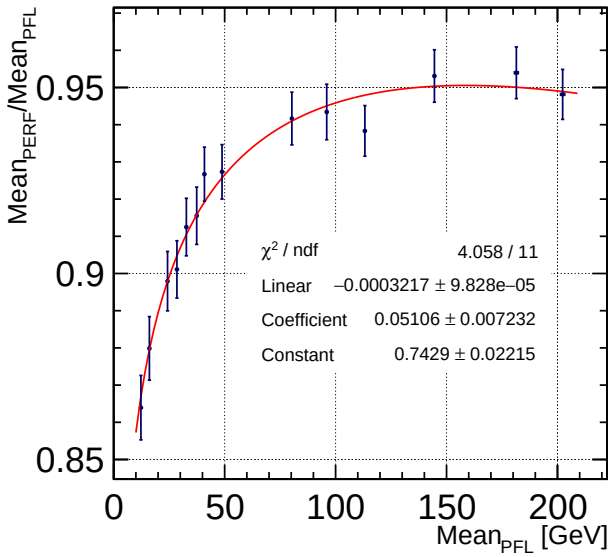


Figure 8.14: The dependency of the energy correction factor, $k = \text{Mean}_{\text{PERF}}/\text{Mean}_{\text{PFL}}$, on the observed neutral energy, determined from fourteen data samples with different $\sqrt{s}$ values. In this case, the observed neutral energy is considered to be the mean value of the $\texttt{SGV}_{\texttt{PFL}}$ neutral energy distribution. The fitted red line represents the resulting scaling function, $k(\text{E}_0^{vis.})$.

The data points were fitted with the function: $k(x) = a \cdot x + b \cdot \ln(x) + c$, where the x variable represents the neutral energy. The obtained parameter values are indicated in figure 8.14.

In order to apply the scaling function on the analysis level, a dedicated processor was written in the `Marlin` framework [159]. Using the reconstruction information stored in the data files, i.e., `PFO` charge and energy, the neutral energy in the

event is computed. The obtained value is then plugged in the scaling function, $k(\mathrm{E}_0^{vis.})$, and the corresponding scaling factor $k_{\mathrm{E}_0^{vis.}}$ is calculated.

The reconstructed momentum and the assumed mass hypothesis are also saved in the data files, for each reconstructed particle. Thus, taking this information into account, the determined energy correction factor $k_{\mathrm{E}_0^{vis.}}$ is then applied to scale the momentum of every neutral PFO in the event.

The neutral energy scaling procedure was applied to the $Z \rightarrow uds$ data samples produced with $\texttt{SGV}_{\texttt{PFL}}$. Its performance was analysed once again in terms of the *total* (charged and neutral) visible energy and the jet energy resolution. The results are summarised in the following.

8.6 Particle Flow Performance of SGV with Confusion Emulation and Neutral Visible Energy Correction

The impact of the neutral visible energy correction on the Particle Flow performance of $\texttt{SGV}_{\texttt{PFL}}$ was investigated. For this purpose, the total visible energy distributions and the jet energy resolutions obtained with $\texttt{SGV}_{\texttt{PFL}}$ *after* the neutral energy scaling (from now on noted as $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$) were compared with the output of: the full simulation (DBD), the fast simulation without the confusion emulation ($\texttt{SGV}_{\texttt{PERF}}$) and $\texttt{SGV}_{\texttt{PFL}}$ *without* the energy correction. Six $Z \rightarrow uds$ data sets, corresponding to the ILC centre-of-mass energies presented in table 8.1 were used in the comparison, for each full and fast simulation variant. The performance of the LoI version of the full simulation was mostly taken into account in the assessment of the jet energy resolution results.

Visible Energy

The comparison in terms of the total visible energy distributions is illustrated in figure 8.15. It was noted that, as intended, the $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ histograms, presented in red, are "shifted" towards lower energy values as a result of the neutral energy correction. Furthermore, their mean values, calculated as Mean_{90}, agree with the full simulation results within 1.2% for all six data samples.

However, it was also observed that the widths of the $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ distributions, expressed as rms_{90}, increased in comparison to their $\texttt{SGV}_{\texttt{PFL}}$ and DBD counterparts. This is particularly noticeable for the higher range of centre-of-mass energies. Thus, for $\sqrt{s}$ values above 360 GeV, the $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ histograms are on average 5% wider than the ones obtained from $\texttt{SGV}_{\texttt{PFL}}$ data.

The discrepancy is even larger when the comparison is made with respect to the DBD distributions: 81% for $\sqrt{s}$=200 GeV and approximately 50% for centre-of-mass energies larger than 360 GeV.

The larger width (rms_{90}) of the $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ visible energy distributions is most likely caused by defining the scaling factor k as the ratio between the fitted *mean* values of the $\texttt{SGV}_{\texttt{PERF}}$ and $\texttt{SGV}_{\texttt{PFL}}$ neutral energy distributions. When applying the

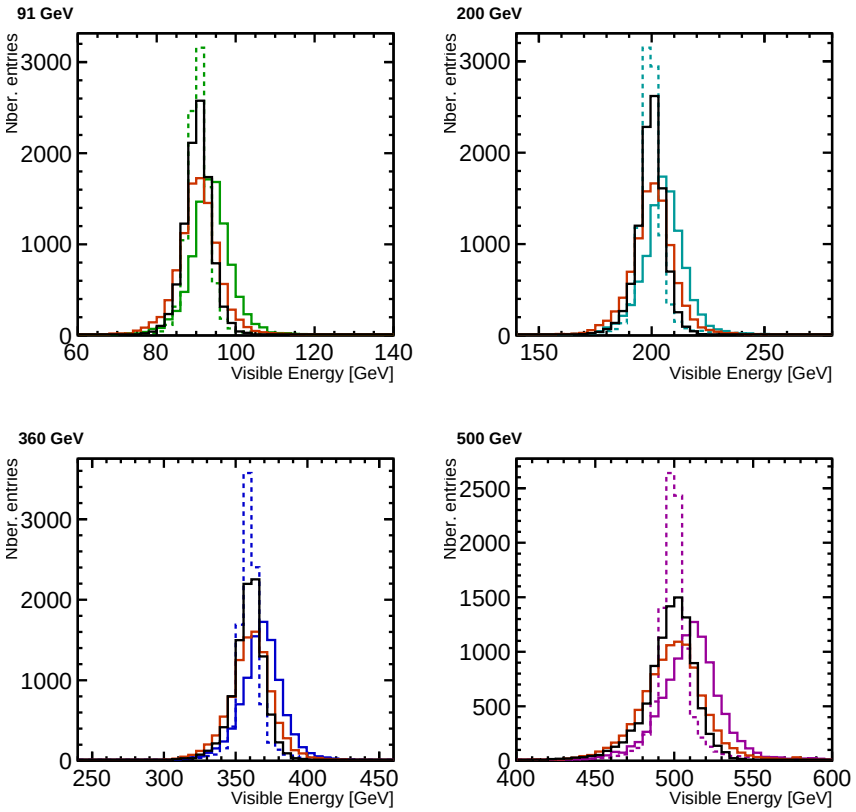


Figure 8.15: The total visible energy distributions for four $Z \to uds$ data sets. The corresponding $\sqrt{s}$ values are indicated in the upper left corners. The $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ distributions, shown in red, were shifted to lower energy values such that their peak position are compatible with the DBD (black line). The $\texttt{SGV}_{\texttt{PERF}}$ and $\texttt{SGV}_{\texttt{PFL}}$ distributions are indicated by the coloured continuous and dashed lines, respectively.

energy correction factor, for events in which the total neutral energy is either smaller or larger than the fitted central value, the calculated k value is, respectively, either too large or too small. Essentially, the procedure scales the entire neutral energy distribution with a factor calculated only with respect to its mean. Consequently, this issue manifests as the observed broadening effect.

Jet Energy Resolution

The jet energy resolution was determined for each of the six $Z \to uds$ data sets that were produced with $\texttt{SGV}_{\texttt{PFL}}$ and then scaled according to the previously discussed procedure. The result is presented (red line and empty red circles) in figure 8.16 as a function of the individual corresponding jet energy.

The increase of the rms_{90} observed for the $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ visible energy distributions has an overall degrading effect on the attained jet energy resolution. The performance obtained with the neutral energy correction is approximately 4% worse than the initial $\texttt{SGV}_{\texttt{PFL}}$ results (light purple) for jet energies below 180 GeV. It degrades further and becomes 9% worse, for jet energies of 250 GeV.

The comparison with the full simulation performance, i.e., taking both the DBD

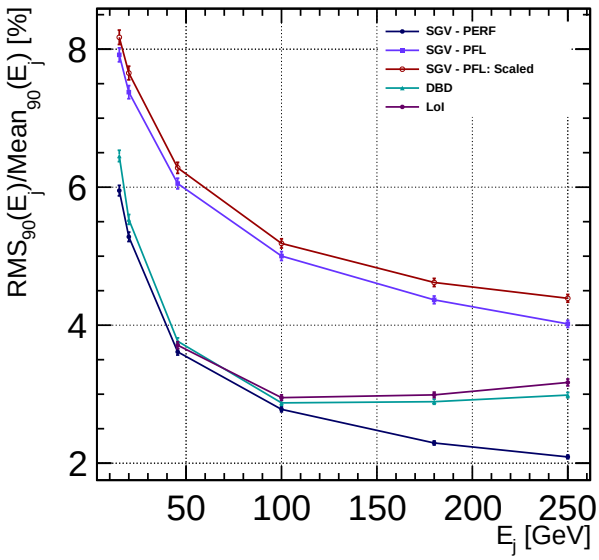


Figure 8.16: The jet energy resolution for all six $Z \to uds$ data samples is shown as a function of the individual jet energy. The performance obtained with $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ is compared to the output of all the other fast and full simulation versions considered in this study.

and the LoI versions into account, reveals a ≈30% deterioration for low energy jets. The largest discrepancy was found for jet energies of 100 GeV where the jet energy resolution attained with $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ is approximately 80% worse than both full simulation versions. This difference is reduced by half, to ≈40% for the highest jet energy.

8.7 Conclusions

The Particle Flow performance of the ILD fast simulation, SGV, was evaluated in terms of the visible energy observed in simulated collision events and the jet energy resolution expressed as a function of the single jet energy.

It was found that, in its default running mode ($\texttt{SGV}_{\texttt{PERF}}$) that does not consider any Particle Flow confusion effects, the fast simulation does not reproduce the performance attained with the ILD full simulation versions LoI and DBD (figure 8.2). This disagreement is especially prominent for higher jet energies, as expected, since the impact of the confusion effects increases with the jet energy. Unexpectedly, an ≈ 8% difference was also found at low jet energies.

In order to alleviate this discrepancy, the behaviour of the Particle Flow reconstruction program, PandoraPFA, used in the context of the LoI reconstruction was studied with respect to several relevant observables. The final goal was that the level of realism provided by SGV is increased by emulating the unavoidable Particle Flow confusion effects.

A first proof-of-principle study performed for this thesis (section 8.3.1) demon-

strated that it is possible to parametrise the behaviour of PandoraPFA in this respect such that the determined parametric functions may be implemented in $\texttt{SGV}_{\texttt{PERF}}$. However, the initial study relied on simplifications that did not fully meet the level of detail requirements of the fast simulation.

Studying the manner in which the Particle Flow confusion can be parametrised and implemented in SGV is currently an ongoing effort. A subsequent, more involved study was performed [6] (section 8.3.2). Its results were implemented in the fast simulation as optional user routines ($\texttt{SGV}_{\texttt{PFL}}$). This represented the first attempt to emulate the confusion effects in the ILD fast simulation.

The evaluation of the $\texttt{SGV}_{\texttt{PFL}}$ performance (section 8.4) revealed that the implemented confusion parametrisations caused a significant deterioration of the jet energy resolution with respect to the full simulation performance. Furthermore, an undesirable bias towards higher values was found in the visible energy.

The overall degradation of the jet energy resolution was classified as a pessimistic estimate. The inadvertent shift introduced in the visible energy was investigated. A dedicated energy correction procedure (section 8.5) was developed to address the bias. It was observed that the realised energy correction method is successful in reducing the initial bias, however, at the cost of deteriorating the jet energy resolution by a few percent. The procedure is intended to be applied on analysis level for all the Monte Carlo data presently produced with $\texttt{SGV}_{\texttt{PFL}}$.

The endeavour to emulate the Particle Flow confusion effects exhibited by PandoraPFA with SGV is continuing. In this sense, the too optimistic track efficiency implementation was clearly identified as an issue that requires further study and improvement. Furthermore, new analysis tools like, for instance, the TrueJet processor [91], are being developed. They enable a more detailed study and a clearer understanding of the inner workings of PandoraPFA with the aim of providing a better confusion emulation.

Presently, the jet energy resolution provided by $\texttt{SGV}_{\texttt{PFL}}$, after applying the energy correction, ($\texttt{SGV}_{\texttt{PFL}}^{Scaled}$) is considered a pessimistic estimation of the detector and Particle Flow reconstruction performances.

This must be regarded in the general context in which most fast simulation programs provide too optimistic results. This is due to the inherent simplifications that the fast simulation programs must implement in order to fulfil the processing time requirements.

It was concluded that, to fully understand the implications of the pessimistic jet energy resolution, the $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ performance must be further investigated and compared to the LoI and DBD results in the context of a more involved physics scenario. For this purpose, a supersymmetry (SUSY) scenario involving the separate pair production of charginos and neutralinos at the ILC was considered.

The chosen processes had already been investigated in the context of the ILD *Letter of Intent* document [109] where they were used to benchmark the ILD and PandoraPFA Particle Flow performance. This informed the decision to use the same physics scenario in the DBD and $\texttt{SGV}_{\texttt{PFL}}^{Scaled}$ evaluation. The details of the SUSY model and its realisation at the ILC are presented in the following section.

CHAPTER 9

Gaugino Pair Production at the ILC

The Particle Flow concept is crucial for reaching the jet energy resolution required for physics studies at the ILC. Two subsequent algorithm implementations, i.e., `PandoraPFA` [7] and `PandoraPFANew` [164], have been developed in the context of the two International Large Detector (ILD) full simulation versions, `LoI` and `DBD`, respectively, described in chapter 6. Furthermore, the unavoidable confusion errors that can occur during the Particle Flow reconstruction have been studied in full simulation and implemented as parametrisations in the fast simulation program `SGV` [6], in order to improve its level of realism, as discussed in chapter 8.

Until now, the performances of both the full and the fast simulation implementations of the Particle Flow algorithm have been evaluated and discussed only in terms of the standardised ILD simulated $Z \to uds$ data. The use of this type of Monte Carlo events provides the advantage of circumventing the application of jet clustering algorithms and enables a direct relation between the ILC centre-of-mass energy and the jet energy.

However, the decays of an off-shell Z boson to light quarks represent only a very narrow and highly specific subset of the interesting interactions that would take place at the ILC. In reality, many final states would be characterised by a large number of hadronic jets that would inevitably require jet clustering and for which the determination of the relative jet energy resolution would no longer be straightforward. Consequently, it is very important to investigate the levels of performance achieved by the different implementations of the Particle Flow approach (i.e., in the full and the fast ILD simulations) in a complete physics scenario.

Moreover, such a study would be especially important in view of the large discrepancies between the jet energy resolution obtained with $\texttt{SGV}_{\texttt{PFL}}$ and the full simulation results, presented in section 8.6. The impact that the pessimistic jet energy resolution provided by $\texttt{SGV}_{\texttt{PFL}}$ may have on the precision with which highly relevant properties like, e.g., particle masses or production cross-sections, are

measured must be evaluated and understood. For this purpose, a supersymmetry (SUSY) physics scenario was chosen.

The current chapter provides a brief overview of the SUSY model considered in the analysis and of the manner in which it could be potentially realised at the ILC. Furthermore, the rationale behind the specific choice of the SUSY model is also presented. Lastly, the current status of the model in view of the most recent LHC SUSY search results is presented.

9.1 The "Point 5" Scenario

The supersymmetry model considered in this thesis was first defined in [195] in the context of the minimal Supergravity model [41], mSUGRA. It was proposed as part of a list of physics processes to be used in benchmarking the performance of the two prospective ILC detectors, ILD and SiD. Due to its position on the list, it became known as the *Point 5* scenario and will be referred to as such throughout this thesis.

The values of the Point 5 parameters are presented in table 9.1:

Parameter	Value
m_0	206 GeV
$m_{1/2}$	293 GeV
$\tan\beta$	10
A_0	0
μ	375 GeV

Table 9.1: Overview of the parameter values assumed in the Point 5 scenario [109].

Considering the mass of the top quark to be $M_t = 178\,\text{GeV}$ and the parameters described above, the mass spectrum calculator `Spheno` [200], [201] provides the following gaugino masses: $M_{\tilde{\chi}_1^0} = 115.7\,\text{GeV}$, $M_{\tilde{\chi}_1^\pm} = 216.5\,\text{GeV}$, $M_{\tilde{\chi}_2^0} = 216.7\,\text{GeV}$, $M_{\tilde{\chi}_3^0} = 380\,\text{GeV}$. The squarks and sleptons are heavier than the gauginos, e.g., $M_{\tilde{q}} \approx$ 600-700 GeV, $M_{\tilde{\tau}_1} = 230.8\,\text{GeV}$ and $M_{\tilde{e}_R} = 237.4\,\text{GeV}$, thus ensuring that $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ decay almost exclusively to the lightest supersymmetric particle (LSP) and the corresponding gauge boson. It must be noted that, since the formulation of the Point 5 scenario predates the current LHC Higgs measurements, the Higgs mass was set to the value of $M_H = 115\,\text{GeV}$. Since the Higgs boson has no direct relevance in the processes studied in this analysis, the pre-LHC benchmark definition was used for compatibility reasons.

The first investigations of the scenario were carried out with the `LoI` version of the full simulation programs developed for the ILD and SiD detectors. The two analyses, which embraced different approaches, as well as their respective results were described in [196] (ILD) and [197] (SiD) and also constituted an important contribution to the *Letter of Intent* documents published by the ILD [109] and SiD [198] scientific communities. The motivation for revisiting the Point 5 scenario in the context of this thesis will be addressed later in this chapter.

9.2 Experimental Aspects and Motivation

The Point 5 analysis was carried out considering the following experimental conditions: the ILC would be operated at a centre-of-mass energy of $\sqrt{s}$=500 GeV with longitudinally polarised e^+e^- beams such that 65% of the positrons in the colliding beam are right-handed while the rest 35% are left-handed and 10% of the electrons in their respective beam are right-handed while 90% are left-handed. Furthermore, it was assumed that, in accordance with the proposed ILC physics programme, a data set corresponding to an integrated luminosity of $\int \mathcal{L}dt$=500 fb^{-1} would be available for the analysis. This assumption was reflected in the amount of produced Monte Carlo data. The same conditions have been considered for the Point 5 studies carried out in the context of this thesis.

The relevant Point 5 processes are the separate pair production of charginos ($e^+e^- \to \tilde{\chi}_1^+\tilde{\chi}_1^+$) and neutralinos ($e^+e^- \to \tilde{\chi}_2^0\tilde{\chi}_2^0$) at the ILC, in the experimental context described above. Both production processes are treated, in turn, as signal channels and as background for each other.

The main goal of the analysis was to determine the level of precision with which the masses of the involved SUSY particles and the production cross-sections for the two signal channels can be measured at the ILC; i.e., in the context of the Particle Flow motivated ILD design [123] and the `PandoraPFA` implementation [7] of the aforementioned reconstruction approach.

Production at the ILC

At e^+e^- coliders with sufficient centre-of-mass energy, like the ILC, the charginos and neutralinos can be produced in pairs both in the s- and in the t-channel as illustrated in the Feynman diagrams below:

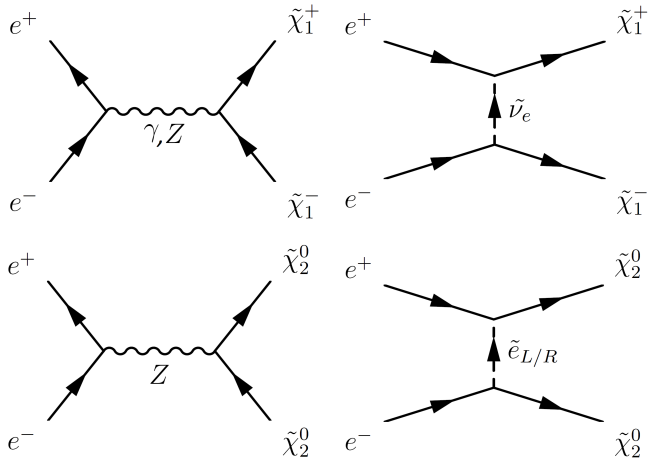


Figure 9.1: Tree level Feynman diagrams depicting the chargino (upper half) and neutralino pair production (lower half) in the $s-$channel (left) and the $t-$channel (right), as it could be realised at the ILC.

The electron and positron helicities determine which production diagrams can dominate.

In the case of annihilation diagrams (i.e., s-channel in figure 9.1), the helicities

of the colliding particles couple directly to each other and, thus, determine the spin of the propagator. In order to be compatible with the Standard Model, the electron and the positron must have opposite helicities, i.e., $e^+_R e^-_L$ or $e^+_L e^-_R$, such that they can annihilate into vector bosons (γ or Z). The other two possible combinations, $e^+_L e^-_L$ or $e^+_R e^-_R$, would require a scalar particle as propagator. One such particle could be the Standard Model Higgs boson. However, its coupling to electrons/positrons is very small, hence, these processes are highly suppressed. Other supersymmetry models that assume R-parity violation, like the one discussed in [202], for instance, allow other scalar particles (e.g., sneutrinos) as mediators. Nevertheless, this is not the case for the R-parity conserving Point 5 scenario discussed in this thesis.

In the case of exchange diagrams (i.e., the t-channel processes from figure 9.1), the helicities of the incoming particles couple directly to the helicities of the final state particles and are, thus, independent from each other. In principle, this would allow all helicity configurations to contribute.

For the chargino pair production, the initial states with identical helicities cannot be realised: for instance, a left-handed positron would require a right-handed sneutrino as propagator while, simulatenously, a left-handed electron would require a left-handed sneutrino. The requirements would be inversed if both incoming particles were right-handed. Furthermore, one of the two cases with opposite helicities, i.e., $e^+_L e^-_L$, can also not be achieved. For this helicity configuration, the common required propagator would be a right-handed (electron) sneutrino ($\tilde{\nu}_{e_R}$). However, this sparticle does not exist in the MSSM since its Standard Model partner is not found in nature. Consequently, the only t-channel diagram that contributes to the chargino pair production is the one which has the $e^+_R e^-_L$ configuration in the initial state and $\tilde{\nu}_{e_L}$ as propagator.

The situation is similar in the case of the t-channel neutralino pair production. The processes involving initial states with identical helicities cannot take place since they would simultaneously require two different propagators (in this case selectrons). The selectron ($\tilde{e}$) is predicted to exist in both helicity states in the MSSM. Consequently, in contrast to the charginos, both configurations with opposite helicities of the initial states, i.e., $e^+_R e^-_L$ and $e^+_L e^-_R$ can be realised.

Decay Modes

Considering the parameters and the calculated sparticle mass values assumed in the Point 5 scenario, the relevant gauginos, $\tilde{\chi}^\pm_1$ and $\tilde{\chi}^0_2$, decay almost exclusively, with branching ratios of $BR = 99.4\%$ and $BR = 96.4\%$, respectively, to the LSP ($\tilde{\chi}^0_1$) and the corresponding gauge boson:

$$\begin{aligned} e^+e^- &\rightarrow \tilde{\chi}^+_1\tilde{\chi}^-_1 \rightarrow W^+\tilde{\chi}^0_1 W^-\tilde{\chi}^0_1 \\ e^+e^- &\rightarrow \tilde{\chi}^0_2\tilde{\chi}^0_2 \rightarrow Z^0\tilde{\chi}^0_1 Z^0\tilde{\chi}^0_1 \end{aligned} \tag{9.1}$$

Since R-parity is conserved in the Point 5 model, $\tilde{\chi}^0_1$ is stable. The LSP interacts only weakly, therefore, from an experimental point of view, the presence of the two LSPs in the final state would manifest as a large amount of missing energy in the observed collision events.

The difference between the chargino mass and the LSP mass is larger than the mass of the W boson: $M_W < |M_{\tilde{\chi}_1^\pm} - M_{\tilde{\chi}_1^0}|$. The situation is similar for the mass difference between the neutralino and the LSP which is larger than the Z boson mass: $M_{Z^0} < |M_{\tilde{\chi}_2^0} - M_{\tilde{\chi}_1^0}|$. Consequently, the produced gauge bosons are real, on-shell particles. They decay mostly to quark-(anti-)quark pairs which immmediately hadronise, thus, manifesting as two hadronic jets for each gauge boson, as illustrated in figure 9.2. The branching ratio for the hadronic decay mode of the W boson is $BR_{W\to qq} = 67.4\%$ [11] and $BR_{Z\to q\bar{q}} = 69.91\%$ [11] for the analogue decay mode of the Z boson.

The gauge bosons can also decay leptonically: $Z^0 \to l\bar{l}$ or $Z^0 \to \nu\bar{\nu}$ and $W^\pm \to l^\pm\nu$, although with lower branching fractions. Furthermore, it is also possible that, since two gauge bosons are produced in a Point 5 signal event, one of them decays hadronically while the other undergoes a leptonic decay.

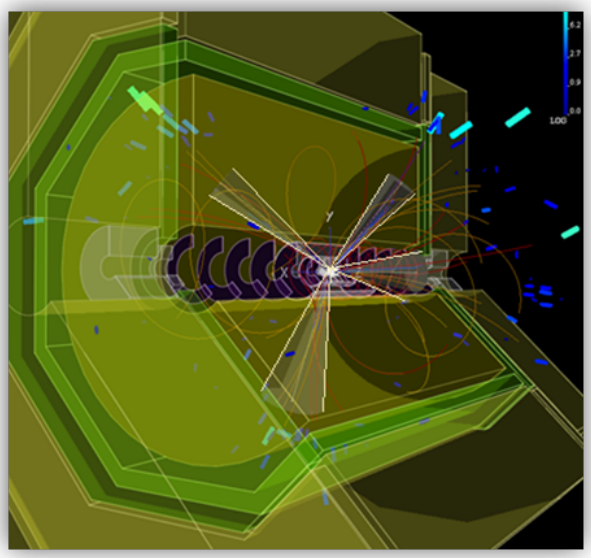


Figure 9.2: Picture from an event display showing the ILD model `ILD_o1_v5` and a $\tilde{\chi}_2^0$ pair production event with the subsequent decay of the emerging Z bosons. The event was simulated with the `DBD` version of the ILD full simulation. The event display uses Monte Carlo truth information to show tracks, clusters and an indication of the boundaries of the four hadronic jets.

From an experimental perspective, the *leptonic* decay modes of the gauge bosons would be easier to identify. Along the large missing energy, caused by the two undetected LSPs, such reactions would leave clear signatures in an otherwise relatively "empty" detector. Thus, the W bosons' decays would be characterised by the presence of only two leptons and extra missing energy due to the associated neutrinos. Similarly, the leptonic decays of the Z bosons would produce either four isolated leptons or only two, depending on whether one of them decays to a pair of neutrinos. Clearly, the cases when both Z bosons decay to neutrinos are not visible in the detector. These signatures would be a clear indication that a Point 5 signal process occurred in the collision event.

In comparison, the *hadronic* decays are characterised by a large particle multiplicity in the final state due to the four hadronic jets, as shown in figure 9.2. This poses a difficult challenge not only for the ILD calorimetric system, but, more importantly, also for the Particle Flow reconstruction. Specifically, when performing the calorimeter cluster formation, `PandoraPFA` is challenged to accurately distinguish between energy deposits produced by different particles in this very busy environment. Furthermore, it must also correctly match the formed clusters with their corresponding tracks, whenever appropriate.

Motivation

As demonstrated in chapter 8, a busy detector environment, caused by a high final state particle multiplicity, inevitably leads to a higher probability for confusion errors to occur in the Particle Flow reconstruction. Consequently, the challenges posed by the hadronic decays in the Point 5 signal processes offer an excellent opportunity to investigate (in a detector simulation) the impact that the Particle Flow performance and the unavoidable confusion would have on the analysis results.

In case of the Point 5 analysis, the most relevant observables are: the reconstructed di-jet energy and the di-jet invariant mass. Their distributions are used to determine the masses of the SUSY particles and the production cross-sections for the two signal processes, respectively. Clearly, these variables are directly affected by the Particle Flow performance and, thus, its effects could be evaluated. These aspects motivated the original study, performed for the ILD *Letter of Intent*, to designate *only* the final states corresponding to the gauge bosons' hadronic decays (equation 9.2) as the signal channels to be studied in the analysis:

$$\begin{aligned} e^+e^- &\rightarrow \tilde{\chi}_1^+\tilde{\chi}_1^- \rightarrow W^+\tilde{\chi}_1^0 W^-\tilde{\chi}_1^0 \rightarrow qq\tilde{\chi}_1^0 qq\tilde{\chi}_1^0 \\ e^+e^- &\rightarrow \tilde{\chi}_2^0\tilde{\chi}_2^0 \rightarrow Z^0\tilde{\chi}_1^0 Z^0\tilde{\chi}_1^0 \rightarrow q\bar{q}\tilde{\chi}_1^0 q\bar{q}\tilde{\chi}_1^0 \end{aligned} \tag{9.2}$$

Furthermore, as can be inferred from the reactions shown in equation 9.2, treating each one of them separately as signal requires the capability to clearly distinguish between the hadronic decays of the W and of the Z bosons, in the context of the large missing energy due to the LSPs. As indicated in chapter 4, the precision with which the hadronic decays of the gauge bosons can be separated is one of the key benchmarks of the Particle Flow performance. In the Point 5 study, it clearly influences the purity of the selected signal samples and, implicitly, also the analysis results. This constituted the main reason for designating the Point 5 physics scenario as a benchmark for evaluating both the ILD calorimetry and `PandoraPFA`'s Particle Flow performance.

One of the central tasks of this thesis concerned the re-evalation of the Particle Flow performance as the `PandoraPFA` software evolved from its `LoI` implementation to the `PandoraPFANew` version, employed in the `DBD` full simulation. Furthermore, another main objective was to determine the impact that the significantly more conservative jet energy resolution conferred by the fast simulation $\texttt{SGV}_{\texttt{PFL}}$ (chapter 8) may have on the precision with which the relevant observables can be

determined in an involved physics analysis. For these purposes and in view of the experimental aspects discussed above, the Point 5 scenario was considered once again a well suitable study case.

9.3 Current Status of the LHC Searches

The supersymmetry (SUSY) searches carried out by the two general purpose experiments at the Large Hadron Collider (LHC), ATLAS [203] and CMS [204], are confronted with the challenge of a vast (model) parameter space. In order to interpret the experimental results and set discovery limits on various supersymmetric scenarios, the analyses rely on *simplified models* [205–207] in which the masses of the relevant SUSY particles are the only free parameters.

The simplified model relevant for the Point 5 scenario similarly assumes that $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ are both wino-like and mass degenerate while $\tilde{\chi}_1^0$ is bino-like and lighter. All the other sparticles are considered to be heavier and decoupled. Both ATLAS [208] and CMS [209] experiments employed the same simplified model for their data interpretation.

In contrast to the Point 5 analysis, which was elaborated as a benchmark for the Particle Flow performance, both LHC searches focus on a different production channel, more advantageous in the context of the experimental environment at the LHC. The production channel and its two potential final states are illustrated in figure 9.3. In this case, both the hadronic (left) and the leptonic (right) decays of the on-shell W boson are taken into account while only the leptonic decay mode of the Z boson is considered.

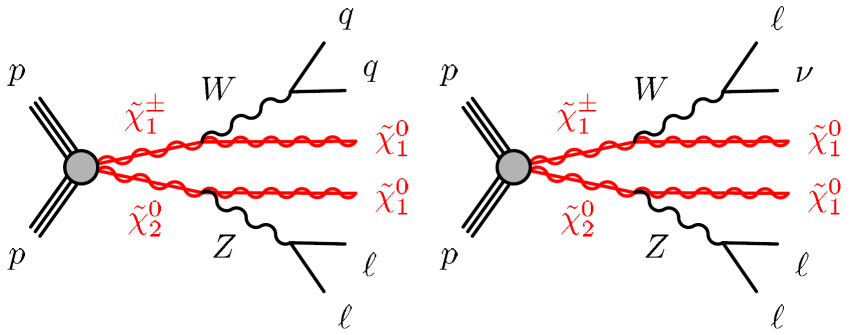


Figure 9.3: Feynman diagrams illustrating the two potential final states of the signal process $pp \to \tilde{\chi}_1^\pm \tilde{\chi}_2^0$ considered in the relevant LHC SUSY searches. Figure taken from [208].

As in all simplified models, it was assumed that the gauginos decay to the LSP and the corresponding gauge boson with branching ratios of $BR_{\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 W^\pm} = BR_{\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 Z^0} = 100\%$. The typical Standard Model branching ratios are considered for the decay modes of the W and Z gauge bosons.

The designation of the mixed chargino-neutralino ($\tilde{\chi}_1^\pm \tilde{\chi}_2^0$) channel as search signal, i.e., instead of the direct pair production processes ($\tilde{\chi}_1^+ \tilde{\chi}_1^-$ and $\tilde{\chi}_2^0 \tilde{\chi}_2^0$) considered in the Point 5 study, is motivated by the higher cross-section of the former.

For illustration, using the Point 5 parameter values and sparticle masses, the cross-sections of the relevant processes were computed with the `MadGraph` [210] program, assuming the LHC being operated at a $\sqrt{s}$=13 TeV centre-of-mass energy. The results [211] showed that $\sigma(pp \to \tilde{\chi}_1^{\pm}\tilde{\chi}_2^0)$ is a factor of two larger than $\sigma(pp \to \tilde{\chi}_1^+\tilde{\chi}_1^-)$ and up to five orders of magnitude larger than $\sigma(pp \to \tilde{\chi}_2^0\tilde{\chi}_2^0)$.

Furthermore, considering the large QCD background and the unavoidable pile-up at the LHC, the signal final states should enable a straightforward identification. In this context both final states shown in figure 9.3 are potentially good candidates. The one consisting of three isolated leptons and a large amount of missing energy also provides the advantage of a rather small Standard Model background [212].

In contrast, the more complex topology of the final states treated as signal in the ILC Point 5 analysis, i.e., comprising four hadronic jets, would constitute an unnecesary complication for the signal identification at the LHC.

A detailed description of the most recent searches performed by the ATLAS and CMS experiments, concerning the simplified model mentioned above, including the definition of the signal regions and the background treatment, can be found in [208] and [209], respectively. Both analyses rely on recorded data samples corresponding to approximately 36 fb^{-1} of proton-proton collisions at the LHC running with a centre-of-mass energy of $\sqrt{s}$=13 TeV. The most relevant results are shown in figure 9.4.

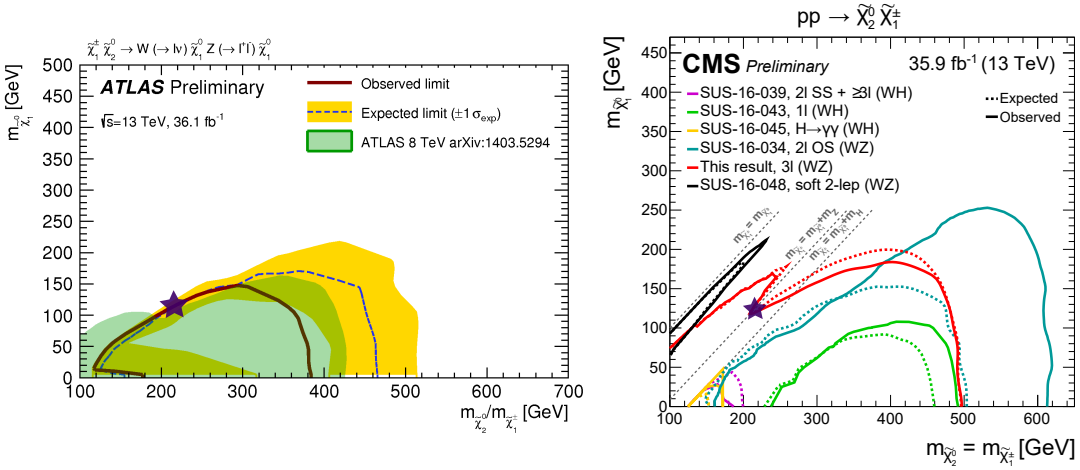


Figure 9.4: The results of the ATLAS (left) and CMS (right) searches for the chargino and neutralino mixed pair production ($\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$), presented as exclusion contours in the mass plane determined by the $\tilde{\chi}_1^0$ (LSP) mass and the degenerate $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ masses. The surface below the curves is the excluded region. The purple star indicates where the Point 5 scenario would be located in the gaugino masses plane. The ATLAS figure was taken from [208] while the CMS result was taken from [209].

The plots in figure 9.4 show the observed (continuous line) and expected (dashed line) exclusion limits at the 95% confidence level on the $\tilde{\chi}_1^0$ for degenerate $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ masses based on the three lepton final state. The ATLAS figure (left) also shows, in yellow, around the expected limit, the $\pm 1\sigma$ contour band

containing all uncertainties apart from the theoretical uncertainties on the signal cross-section.

The CMS counterpart is represented by the red line from the right pannel of figure 9.4. In addition, the CMS plot also illustrates with the light blue line the exclusion limits determined when searching for the other relevant final state, comprising the hadronic decays of the W boson.

The purple star symbol in both panels of figure 9.4 indicates where the Point 5 scenario would be located. It can be seen that, in light of the most recent LHC results, the model is severely under pressure. To obtain a more quantitative evaluation of the status of the Point 5 scenario an analysis with the `CheckMATE` program was performed [213].

9.3.1 `CheckMATE` Analysis

The `CheckMATE` program [214, 215] was specifically designed for testing a large variety of SUSY models in terms of their validity with respect to the published results from the most recent LHC searches.

It can take as input a SUSY model description file written in the Les Houches Accord (SLHA) format [150]. Based on this information, Monte Carlo events are generated with `MadGraph` [210].

In the case of the Point 5 evaluation, the same SLHA file that was used for the ILC Monte Carlo event generation (with `Whizard 1.95`) was provided with one exception: the squark masses were increased by a factor of ten with respect to their initial Point 5 values. This was done in order to overcome the exclusion limits placed on squark masses by the ATLAS results presented in [220]. The mentioned analysis excludes squark masses up to approximately 700 GeV. Nevertheless, since in the Point 5 scenario the squark sector is completely decoupled from the relevant gauginos, this modification can be performed.

Using this data, 100000 events were generated for each relevant signal process, considering the LHC running at a centre-of-mass energy of $\sqrt{s}$=13 TeV. The event numbers were then normalised to the respective process cross-sections.

The generated events are then typically processed through a fast simulation of the ATLAS detector[1] [97]. The detector simulation is based on a well tuned implementation of `Delphes 3` [170]. The acceptance region of the ATLAS detector, the trigger as well as the various reconstruction and identification efficiencies are modelled in the simulation. The outcome is a set of final state objects for each Monte Carlo event.

A large number of so-called "cut-and-count" ATLAS supersymmetry searches, specifically, their experimental selection cuts and signal region definitions, are implemented in `CheckMATE`. For each analysis, the program applies the *exact same* cuts on the newly simulated events (i.e., in this case, Point 5 events). The outcome consists of a number of signal events, S_a, that survive the specific cuts

[1] In `CheckMATE 2` [215], the CMS detector is also implemented. However, the relevant CMS SUSY analyses were not yet implemented at the time of writing. Consequently the evaluation of the Point 5 model point was performed only with respect to the ATLAS results.

of analysis "a" and the uncertainty, ΔS_a, on that number. ΔS_a includes both the statistical uncertainty coming from the (limited) number of produced Monte Carlo events as well as the systematic uncertainty estimated from the signal cross-section uncertainty.

Each ATLAS SUSY search, a, included in the comparison provides: *(i)* the number of observed (real) events, S_a^{obs}, in the context of its specific event selection and *(ii)* the number of events, S_a^{SM}, given by the Standard Model prediction, considering the same experimental cirumstances.

Using these two event numbers it is possible to determine, model-independently, the maximum number of new signal events, N_a^{max} that could be added to the Standard Model prediction, S_a^{SM}, without violating the experimental observation, S_a^{obs}, by more than two standard deviations. The choice of 2σ is motivated by the fact that it approximately corresponds to the 95% confidence level typically used in the LHC results formulation [214](illustrated for example in the left plot shown in figure 9.4). `CheckMate` calculates N_a^{max} using profile likelihood functions [214].

In order to verify the compatibility of the SUSY model with the ATLAS experimental results, the obtained number of surviving signal events, S_a, must be compared to N_a^{max}. This comparison is performed for each ATLAS search included in the `CheckMate` analysis. For this purpose, the parameter r_a is defined [215]:

$$r_a \equiv \frac{S_a - 1.64 \times \Delta S_a}{N_a^{max}} \tag{9.3}$$

The numerator represents the lower limit of the 95% confidence interval centred on the number of signal events determined by `CheckMate` for a specific ATLAS analysis. Consequently, a SUSY model is considered to be excluded if any computed value of the r_a parameter is larger than one.

The Point 5 scenario was analysed [213] with respect to 18 different ATLAS SUSY searches. The most relevant results are summarised in table 9.2.

Analysis (a)	Signal Region	S_a	ΔS_a	N_a^{max}	r_a
ATLAS_1604_07773 [216]	IM4	3.86	0.33	491	0.007
ATLAS_1605_03814 [217]	2jm	0.72	0.14	48	0.010
ATLAS_CONF_2016_054 [218]	GG2J	0.87	0.32	20	0.017
ATLAS_CONF_2016_076 [219]	SR_w_sf	2.32	0.48	11.1	0.138
ATLAS_CONF_2017_022 [220]	3j-1300	3.78	1.04	86	0.024
ATLAS_CONF_2017_039 [208]	SR3-WZ-0Ja	11.32	2.00	11.35	0.708
ATLAS_CONF_2017_040 [221]	ee	9.75	1.46	104	0.071

Table 9.2: Overview of the `CheckMATE` [215] results [213] concerning the compatibility of the Point 5 scenario with the current exclusion limits determined from observed data by the ATLAS experiment. The numbers of events are given in fractional form since they were normalised to the cross-sections of the respective production processes.

It can be seen from table 9.2 that the most sensitive ATLAS analysis is, as expected, the one presented in [208] which was also discussed earlier in this section.

The most relevant signal region corresponds to the search for a WZ topology in the final state. The left plot in figure 9.4 corresponds to this signal region.

`CheckMATE` predicts that, in the context of this specific analysis, the Point 5 scenario would produce $S_a = 11.32$ signal events with an uncertainty of ΔS_a=2 events. The calculated maximum number of excess events still compatible with the ATLAS experimental observation is N_a^{max}=11.35. Consequently, the Point 5 model is barely still allowed. This is also illustrated by the corresponding value of the r parameter, r_a=0.708. However, it is very likely that the model will be excluded by the following iteration of this ATLAS analyis or by its CMS counterpart.

CHAPTER 10

Analysis Strategy

The supersymmetry scenario known as Point 5 [109, 195] was chosen as a study case for the Particle Flow performance investigations that constitute the main focus of this thesis. The present chapter provides an overview of the data samples used in the analysis. Furthermore, in section 10.2, the most relevant Standard Model background processes are indicated and the strategies used for performing the signal selection are described. The signal selection procedure follows the one originally applied in the ILD *Letter of Intent* study [196] to a great degree. However, a few improvements were made in view of the increased level of realism of the `DBD` full simulation version. They are also discussed in the following.

10.1 Data Samples

In order to carry out the performance assesment of the three different realisations of the *Particle Flow* algorithm, corresponding to the `PandoraPFA` (`LoI`), `PandoraPFANew` (`DBD`) and $\texttt{SGV}_{\texttt{PFL}}$ implementations, three different sets of Monte Carlo data samples were employed.

Data Samples Nomenclature

The general nomenclature for the three sets of data samples used throughout this thesis is described in the following:

- The LOI sample: The Monte Carlo data contained in this set has been mass produced with the `LoI` version of the ILD full simulation. It was originally used in the analyses presented in the ILD *Letter of Intent* [109].
- The DBD sample: It comprises Monte Carlo data that was mass produced with the `DBD` version of the ILD full simulation. The samples were employed in the studies carried out and published in the ILC *Technincal Design Report* [123].

- The $\mathsf{SGV}_{\mathsf{PFL}}$ sample: It consists of events that were processed with the fast simulation program `SGV` [6] run with the implemented Particle Flow confusion emulation.

The main differences between the `LoI` and `DBD` versions of the ILD full simulation and, implicitly, the PandoraPFA-based [7] full Particle Flow reconstruction versions have been discussed in chapter 6. The working principles of the `SGV` fast simulation have been presented in chapter 7. The Particle Flow confusion emulation implemented in $\mathtt{SGV}_{\mathtt{PFL}}$ was described in chapter 8. An overview of the most important differences between the three data sets used in this thesis is presented in table 10.1.

In view of the comparisons that will be performed on the analysis level, all three Monte Carlo data sets contain the same physics processes. Based on their final states, they can be categorised and grouped into different classes. A summary of the reactions considered in the following studies and their classification are presented in table 10.2.

Monte Carlo Data Production Aspects

In addition, there were also several technical details concerning the data generation that must be noted. They are presented in the following.

→ **Shift of nominal W boson mass value in signal samples**: During the generation of the signal processes, the mass of the W boson was lowered by `Whizard` to the value of $M_W = 79.8$ GeV, as shown in figure 10.1. This effect is found in the SUSY signal samples of all three (i.e., LOI, $\mathsf{SGV}_{\mathsf{PFL}}$ and DBD) data sets.

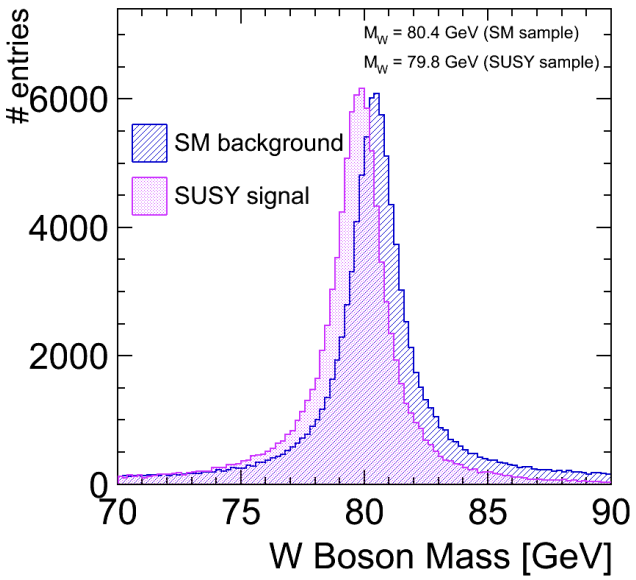


Figure 10.1: Comparison of the W boson mass, M_W, in the Standard Model background sample and in the Point 5 chargino signal sample. Due to an internal consistency check in `Whizard`, M_W is shifted by the event generator to the lower value of 79.8 GeV.

		LoI	DBD	SGV_{PFL}
Gen. Level	Whizard version	v01.40	v01.95	v01.95
	Beamspectrum	RDR [54]	TDR [57]	TDR [57]
	PYTHIA tune	Default [154]	OPAL [153]	OPAL [153]
Sim. & Reco. Level	ILD model	ILD_00	ILD_o1_v5	sgv_geo3
	Tracking	Fortran, LEP-based	C++ new implement.	inverse Kalman filter
	Particle Flow	PandoraPFA	PandoraPFANew	Confusion param.
	$\gamma\gamma$ overlay	No	Applied	No

Table 10.1: Overview of the most relevant differences regarding the production of the three data sets used in this thesis.

Class Name	Processes	Type
$\tilde{\chi}_1^\pm$-hadronic	$\tilde{\chi}_1^\pm\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 qq\tilde{\chi}_1^0 qq$	Signal
$\tilde{\chi}_2^0$-hadronic	$\tilde{\chi}_2^0\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 qq\tilde{\chi}_1^0 qq$	Signal
SUSY other	$\tilde{\chi}_1^\pm\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 qq\tilde{\chi}_1^0 l\nu$ $\tilde{\chi}_1^\pm\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 qq\tilde{\chi}_1^0 l\nu$ $\tilde{l}^+\tilde{l}^- \to l^+l^-$ $(l{=}e,\mu,\tau)$	Background
SM 2f	$Z^0 \to q\bar{q}$ $Z^0 \to ll$	Background
SM 4f	$W^+W^- \to qqqq$ $W^+W^- \to qql\nu$ $W^+W^- \to l\nu l\nu$ $W^\pm Z^0 \to qqqq$ $W^\pm Z^0 \to l\nu ll$ $Z^0Z^0 \to qqqq$ $Z^0Z^0 \to qqll$ $Z^0Z^0 \to llll$	Background
SM 6f	$t\bar{t}, llW^+W^-, qqW^+W^-, l\nu W^+W^-, qqqqZ$	Background

Table 10.2: Overview of the physics processes and of their classification employed in the Point 5 studies presented in this thesis. Since the e^+e^- interaction is clearly implied, only the intermediary and/or final states are shown.

This issue stems from the choice of the electroweak renormalisation scheme, as explained in the following. The physics model that is employed as input to `Whizard` must be defined according to the standard set by the **SUSY Les Houches Accord** (SLHA) [150]. The convention employed by SLHA is that the electroweak sector is determined by:

- M_Z: the Z boson pole mass, $M_Z = 91.1876 \pm 0.0021$ GeV [11]
- G_F: the Fermi constant, $G_F = 1.166364 \times 10^{-5}$ GeV^{-2}, determined from the expression of the muon lifetime
- $\alpha_{em}(M_Z)^{\overline{MS}}$: the electromagnetic coupling at the Z pole in the minimal subtraction ($\overline{\text{MS}}$) scheme

From these inputs, other electroweak parameters such as $\sin^2\theta_W$ and M_W are determined in `Whizard`. The value of the former depends on the chosen renormalisation scheme. In the software setup for event generation, the chosen renormalisation scheme was the *on-shell* scheme, under which $\sin^2\theta_W$ is defined as $\sin^2\theta_W = 1 - M_W^2/M_Z^2$ for all orders in perturbation theory. Therefore, $M_W = M_Z cos\theta_W$. In order to conserve the electroweak gauge invariance, this relation must be fulfilled and, thus, the W boson mass was shifted by `Whizard` from its correct value to M_W=79.8 GeV.

→ **Different beam energy spectra**: Two different sets of ILC beam parameter values were used in the subsequent Monte Carlo data productions, as presented in section 6.2 and indicated in table 10.1. Consequently, two different beam energy spectra were determined based on the aforementioned parameters. The generation of Monte Carlo events considering different beam energy spectra led to differences in the obtained cross-sections. For instance, the cross-sections for the Point 5 signal processes calculated by `Whizard` while taking the TDR parameters into account are on average 3.56 % smaller than the ones calculated with the RDR parameters. A detailed list of the considered SUSY and Standard Model processes and of the differences in their respective cross-section values is presented in table A.1 shown as an appendix in this thesis.

→ **Missing signal processes**: Due to a technical issue in the `v01.95` version of the `Whizard` event generator some neutralino decay channels, specifically those containing four quarks with the same flavour in the final state (e.g., $e^+e^- \to \tilde{\chi}_2^0\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 b\bar{b}\tilde{\chi}_1^0 b\bar{b}$), could not be generated. A possible solution requires further investigation. The problem affects a number of events smaller than 50. It manifests as a reduction of 0.6% of the total number of events in the neutralino signal sample. This issue concerns only the DBD and SGV_{PFL} data sets.

Data Weighting

The data weighting procedure is identical for all three data sets, i.e., LOI, DBD and SGV_{PFL}, therefore, it will be discussed as a general principle in the following.

One of the notable advantages of the International Linear Collider (ILC) is the capability to provide longitudinally polarised electron and positron beams. The (longitudinal) beam polarisation is defined as:

$$\mathcal{P} = \frac{N_R - N_L}{N_R + N_L} \tag{10.1}$$

where N_R and N_L represent the percentage of particles in the beam with right-handed and left-handed helicity, respectively.

The processes presented in table 10.2 were simulated for all possible (allowed) helicity combinations of the colliding positron and electron: $h(e^+e^-) \in \{LL, LR, RL, RR\}$. Moreover, it was considered that each incoming beam is 100% polarised.

However, for the Point 5 physics scenario that forms the focus of this thesis a different polarisation configuration of $\mathcal{P}(e^+, e^-)$=(+30%, -80%) was assumed. In this context, the minus sign typically denotes left-handedness while the plus sign indicates right-handedness, respectively. In view of equation 10.1, a positron beam polarisation of $\mathcal{P}_{e^+}$=(+30%) means that 65% of the particles in the positron beam are right-handed and 35% are left-handed, i.e., $N_R = 0.65$ and $N_L = 0.35$. Analogously, for a $\mathcal{P}_{e^-}$=(-80%) polarised electron beam, $N_R = 0.1$ and $N_L = 0.9$.

In order to reconcile the Monte Carlo data with the specific polarisation configuration assumed in the Point 5 study, a weighting factor must be applied to all samples, for each helicity combination. Thus, four different weighting factors are determined:

$$\begin{aligned} w^{pol}(e^+_R, e^-_L) &: 0.35 \times 0.1 = 0.035 \\ w^{pol}(e^+_L, e^-_R) &: 0.65 \times 0.9 = 0.585 \\ w^{pol}(e^+_R, e^-_R) &: 0.35 \times 0.9 = 0.315 \\ w^{pol}(e^+_L, e^-_L) &: 0.65 \times 0.1 = 0.065 \end{aligned} \tag{10.2}$$

Furthermore, in the Point 5 study, it was assumed that both the Standard Model and the SUSY data samples correspond to an integrated luminosity of $\int \mathcal{L}dt$=500 fb^{-1}. However, while the Monte Carlo production of the SUSY samples does indeed correspond to the required integrated luminosity, this is not always the case for the Standard Model background data. Due to the limited available computing resources some SM samples are smaller. In these cases, a different (luminosity) weight must be applied such that, in the end, a consistent data set is available for the physics analysis. The luminosity weight w^{lumi} is defined as:

$$w^{lumi}_{proc,h} = \frac{\sigma_{proc,h} \times \int L\mathrm{d}t}{N_{proc,h}} \tag{10.3}$$

where $\sigma_{proc,h}$ is the production cross-section of process *proc* for a helicity combination h and $N_{proc,h}$ is the number of events produced for that specific process, with the respective helicity combination.

Consequently, the final weighting factor that accounts for both the assumed polarisation and the integrated luminosity is: $w_{proc,h} = w^{pol}_h \cdot w^{lumi}_{proc,h}$

This weight is calculated for each process and helicity combination. It is then applied on an event-by-event basis. Thus, after weighting, the final event numbers are normalised such that they correspond to the designed integrated luminosity and polarisation configuration.

10.2 Jet Clustering and $\gamma\gamma$ Background Removal

As described in section 9.2, the relevant (SUSY) signal final states are characterised by the presence of four hadronic jets coming from the decays of the two gauge bosons and a large amount of missing energy due to the two escaping lightest supersymmetric particles (LSPs).

Consequently, the first step in processing the data samples involves grouping the reconstructed particles (`PFO`s) present in the event into jets. The fundamental working principles of the two main types of jet clustering algorithms were briefly presented in section 4.1.1.

Since the pursued signal final states consist of four hadronic jets, all SUSY and Standard Model data samples are "forced" into this configuration. Specifically, the applied jet algorithm is strictly required to form exactly four jets even when the true event topology could be different.

The same algorithm is used for all three Monte Carlo data sets, namely, a variant of the k_T jet clustering algorithm [102] known as the *Durham algorithm*. It is summarily described in the following.

The Durham Algorithm

Developed especially for e^+e^- collisions, the Durham algorithm is an *inclusive* sequential recombination algorithm. It computes a single distance between the `PFO`s inscribed on the input list, namely:

$$dist_{ij}^{Durham} = \frac{2min(E_i^2, E_j^2) \cdot (1 - cos\theta_{ij})}{E_{visible}^2} \tag{10.4}$$

where E_{ij} represents the energy of a pair of reconstructed particles, while θ_{ij} is the angle between them.

The two `PFO`s with the smallest $dist_{ij}^{Durham}$ are combined by summing up their four-momenta, thus, creating a `pseudojet` object. In the Durham algorithm, there is no pre-defined stop condition, thus the procedure continues until *all* reconstructed particles have been grouped into the required number of jets. In contrast to the exclusive k_T algorithm, which will be described next, the input `PFO`s are not tested for being beam particles and therefore no particle is discarded.

The Durham algorithm also provides two variables that can be used to estimate the quality of the clustering. Thus, $y_{n,n+1}$ represents the distance between two reconstructed particles or two pseudojets before being merged while $y_{n-1,n}$ expresses the minimum distance between the existing pseudojets. These values are constantly updated throughout the clustering process.

In a successful and stable jet configuration, the distance between two `PFO`s or protojets which were merged in the last step, $(y_{n,n+1})$, should be very small. Con-

versely, the distance between two (pseudo)jets ($y_{n-1,n}$) should be large. These two output parameters allow the event selection to be refined such that only those events that exhibit the most stable four-jet configuration are selected as potential signal events.

$\gamma\gamma$ Background Removal

In contrast to the LOI and SGV_{PFL} data sets, the events in the DBD samples are overlayed with the $\gamma\gamma \rightarrow hadrons$ background presented in section 3.3.2.

In order to remove the hadrons produced in the beam background $\gamma\gamma$ interactions, an exclusive k_T algorithm (also known as the *longitudinally invariant exclusive k_T jet algorithm* [184]) is applied exclusively on the DBD data set.

The distance measure employed in the exclusive longitudinally invariant k_T algorithm is given by:

$$dist_{ij} = \frac{min(p_{Ti}^2, p_{Tj}^2) \cdot \Delta R_{ij}^2}{R^2} \tag{10.5}$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with $y_{i,j}$ being the rapidities, $\phi_{i,j}$ the azimuthal angles and $p_{Ti,j}$ the transverse momenta of the particles forming the pair. Here, R can be interpreted as a measure for the jet radius and it is a tunable parameter. The number of jets that the algorithm is required to constitute is also a free input parameter.

The above-mentioned pair-distance is calculated for each pair of reconstructed particles. In addition, another type of distance is computed for each individual PFO, namely the distance between the particle and the beam (beam-distance): $dist_{Beam}^i$ which is defined as p_{Ti}^2.

In the next step, the minimum between all the pair-distances ($dist_{ij}$) and the beam-distances ($dist_{Beam}^i$) is calculated. If the calculated minimum is a pair-distance then the two PFOs forming the pair are merged by adding their four momenta. This procedure is known as the *E-scheme* recombination method. Conversely, in case a beam-distance is the found minimum then that particular PFO is considered to be part of the beam and is discarded. Subsequently, all the previous steps are repeated until the number of remaining `pseudojets` (objects formed by repeatedly merging PFOs) matches precisely the requested number of jets.

The hadrons produced in the $\gamma\gamma$ background interactions are emitted close to the beam direction and thus they are expected to fulfill the algorithm's $dist_{Beam}^i \rightarrow min$ rejection condition. This renders the exclusive k_T clustering algorithm a useful tool for removing this type of background.

Since both the number of required jets and the R-value are free parameters of the exclusive k_T algorithm, they can be tuned in order to optimise the clustering performance and the $\gamma\gamma$ background removal. This has been done in this thesis by studying a Monte Carlo sample of $\approx$ 40000 pure $\chi_1^{\pm}$ events.

Three different jet configurations have been used with three different R values for each jet configuration, as shown in Table 10.3. Increasing the number of required jets can be viewed as similar to increasing the "resolution" with which the recombination is performed; thus, the higher the number of jets the smaller the investigated R value.

Jets nr.	4 jets	6 jets	8 jets
R values	1.1	0.9	0.8
	1.3	1.1	1.0
	1.5	1.3	1.2

Table 10.3: Tested jet configurations for the R parameter optimisation.

Figure 10.2 shows a comparison between three jet number and R-value combinations that were considered to be best, namely: 4 jets with R = 1.3, 6 jets with R = 1.1 and 8 jets with R = 1.0. The main observables, i.e., the reconstructed W boson mass (a) and energy (b), have been plotted for each of these configurations. In addition, the Durham algorithm is also plotted for comparison.

Furthermore, to showcase the impact of the $\gamma\gamma$ background, the Durham algorithm has been applied both over a sample with as well as over one without $\gamma\gamma \rightarrow$ hadrons events.

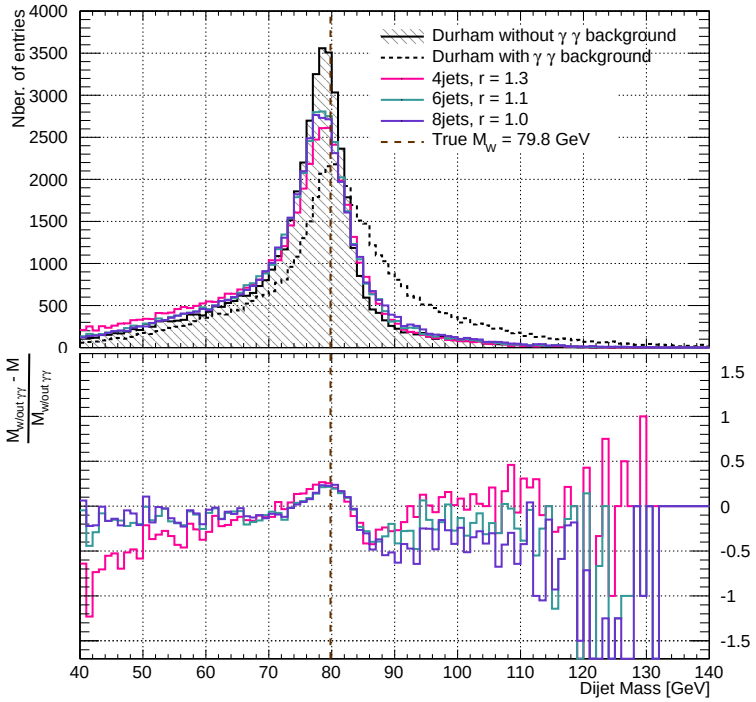
The main reason for choosing the Durham algorithm as reference is that it has consistently provided one of the best performances in the context of e^+e^- events. Furthermore, it does not take the jet radius as an input parameter. Thus, the comparison between the three jet configurations is performed with respect to the (R-)independent Durham algorithm.

It is important to note that, after applying the described exclusive k_T algorithm for $\gamma\gamma$ background removal, one would ideally obtain a distribution very similar to the Durham no background case (solid black). Therefore, the task is to determine which jet number and R-value configuration provides the closest distributions to the no background ones.

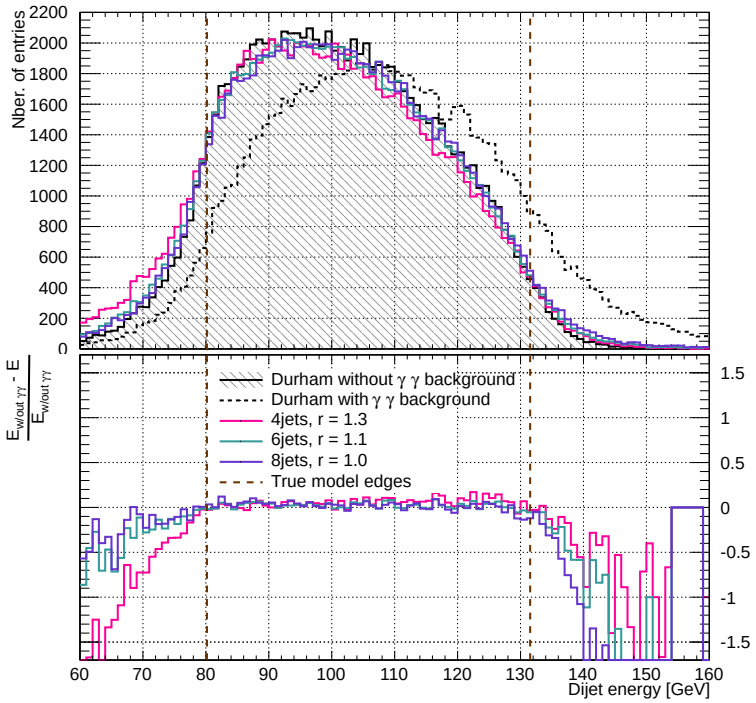
For this purpose, a χ^2 test has been performed comparing each jet configuration to the distribution without $\gamma\gamma \rightarrow hadrons$. Table 10.4 shows the χ^2 divided by the number of degrees of freedom for each comparison, both for the W boson mass as well as for the W energy distributions.

Nr. Jets	R value	χ^2/ndf M_W	χ^2/ndf E_W
4 jets	1.1	24.2	28.5
	1.3	13.4	11.6
	1.5	15.6	7.6
6 jets	0.9	15.6	11.7
	1.1	6.9	4.7
	1.3	13.5	8.2
8 jets	0.8	16.9	10.8
	1.0	9.3	6.8
	1.2	17.7	13.4

Table 10.4: χ^2/ndf for all the tested jet configurations. The combination with the smallest χ^2/ndf value is considered the best: in this case, the 6 jets and R = 1.1 configuration.



(a) W boson (di-jet) mass



(b) W boson (di-jet) energy

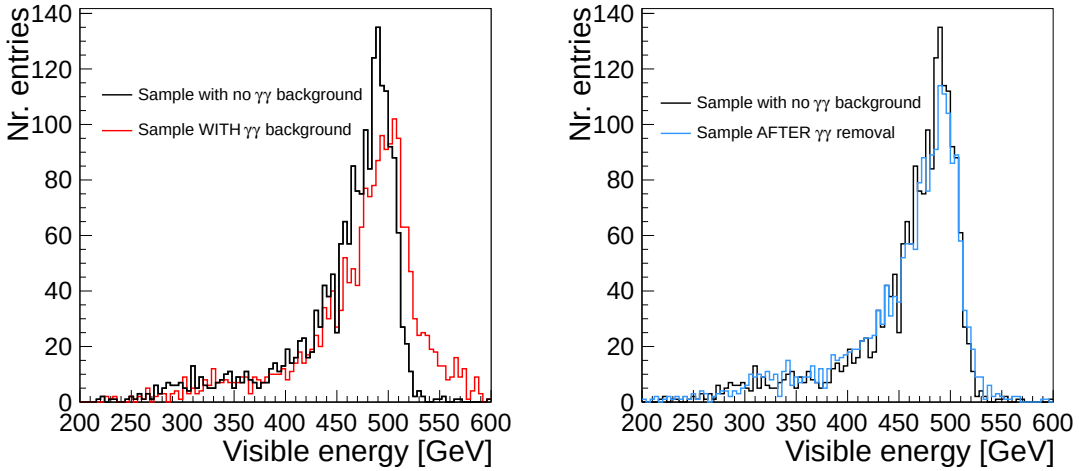
Figure 10.2: Comparison between different jet configurations for the determination of the best R value and jet number.

The lower halves of the two plots in figure 10.2 show the normalised differences between the Durham distribution without $\gamma\gamma$ background and the exclusive k_T distributions after the background removal for the three configurations considered most promising. The best configuration would be the one that provides a difference distribution overall closest to zero and with the smallest fluctuations, especially around the nominal values of the relevant observables (i.e., mass value and energy spectrum edges, marked with vertical dotted lines).

It can be seen from figure 10.2 that the configuration comprising six jets with an individual jet radius of R=1.1 provides the minimum χ^2/ndf value both for the W mass (6.9) as well as in the case of the W energy spectrum (4.7). Thus, this jet configuration was chosen for the $\gamma\gamma$ background removal.

It must be noted that all the physics studies performed on DBD samples and published in the ILC *Technical Design Report* [123] consider only the minimum required number of jets in their jet clustering implementation. This study was the first to demonstrate that requiring a larger number of jets can be beneficial. This is most likely due to the fact that requiring a slightly larger number of jets can account for those cases in which additional jets may be produced when hard gluons are emitted by the quarks or gluons from the final state.

In order to verify the described removal procedure, a Standard Model sample of ≈2000 $WW \rightarrow hadrons$ events has been studied. The outcome is presented in figure 10.3. The sample has been processed: *(i)* without the $\gamma\gamma \rightarrow hadrons$ background (solid black), *(ii)* with the Durham jet algorithm without removing the background (solid red) and *(iii)* with the exclusive k_T algorithm plus, subsequently, with the Durham jet clustering into 4 jets (solid blue).



(a) **(b)**

Figure 10.3: The visible energy in a Standard Model $WW \rightarrow qqqq$ sample study: pannel (a) shows the smearing effect of the $\gamma\gamma$ background on the visible energy while pannel (b) indicates that this effect can be corrected by applying the $\gamma\gamma \rightarrow hadrons$ removal procedure.

It can be seen from figure 10.3 that the $\gamma\gamma$ background removal procedure is

successful in discarding the extra beam-background PFOs that smear the visible energy spectrum. Thus, after applying the aforementioned exclusive k_T algorithm, the visible energy is very similar to the no background case.

10.3 Kinematic Fitting

With the aim of improving the resolution of relevant observables like, e.g., the di-jet mass, a kinematic fit is performed in the preliminary stages of the analysis. The following paragraphs provide a brief overview of the kinematic fitting procedure.

Let there be M measured quantities $\vec{y}(y_1, ..., y_M)$, for example, the four-vectors of a jet in the final state of the event. The measured observables may depend on or can be expressed in terms of a number N of parameters $\vec{x}(x_1, ..., x_N)$, such as the azimuthal angle (ϕ) of the jet, etc. Here, N can be larger or equal to M.

Based on the assumed physics scenario, a certain hypothesis can be formulated to describe these particular events. The hypothesis is expressed by a number K of kinematic constraints, e.g., momentum conservation, $\vec{f_k}(x_1, ..., x_N) = 0$. The measured parameters can then be slightly varied within their uncertainties such that the kinematic constraints are fulfilled. This parameter adjustment procedure is called *kinematic fitting*.

The differences between the measured values x_n and the fitted values ξ_n form a χ^2 function. The aim is to find a (global) minimum of the total $\chi^2(\xi_1, ..., \xi_N, x_1, ..., x_N)$ such that the kinematic constraints are satisfied.

A general form of the χ^2 function can be written as:

$$\chi^2 = (\vec{x} - \vec{\xi})^T \cdot V^{-1} \cdot (\vec{x} - \vec{\xi}) \tag{10.6}$$

where V^{-1} represents the covariance matrix of the measured parameters.

In order to implement the constraints, the method of Lagrange multipliers [222] is used and, thus, an additional contribution χ^2_T can be defined as:

$$\chi^2_T(\xi_1, ..., \xi_N, \lambda_1, .., \lambda_k, x_1, ..., x_N) \tag{10.7}$$

where $\vec{\lambda} = \lambda_1, .., \lambda_k$ are the Lagrange multipliers. The χ^2_T function is then expressed as a global χ^2 plus a constraint term:

$$\chi^2_T = \chi^2(\xi_1, ..., \xi_N, x_1, ..., x_N) + \sum_{k=1}^{K} \lambda_k \cdot \vec{f_k} \tag{10.8}$$

As mentioned above, the purpose is to find the best set of (fitted) parameters $\vec{\xi}$ that minimise χ^2. Thus, the first derivatives in equation 10.8 must be considered and, by setting them to zero, a set of equations corresponding to the minimum conditions can be written:

$$\frac{\partial \chi^2_T}{\partial \xi_m} = \frac{\partial \chi^2}{\partial \xi_n} + \sum_{k=1}^{K} \lambda_k \cdot \frac{\partial \vec{f_k}}{\partial \xi_n} = 0 \tag{10.9}$$

$$\frac{\partial \chi^2_T}{\partial \lambda_k} = \vec{f_k} = 0 \tag{10.10}$$

Equation 10.10 expresses the kinematic constraints. The following matrix notations are useful to be introduced:

$$\vec{X} = (\xi_1, ..., \xi_N, \lambda_1, .., \lambda_k)^T \tag{10.11}$$

$$\vec{Y}(\vec{X}) = (\frac{\partial \chi^2_T}{\partial \xi_1}, .., \frac{\partial \chi^2_T}{\partial \xi_N}, \frac{\partial \chi^2_T}{\partial \lambda_1}, .., \frac{\partial \chi^2_T}{\partial \lambda_K})^T \tag{10.12}$$

The minimum conditions can be then written as $\vec{Y}(\vec{X}) = 0$. This system of equations is then solved in an iterative procedure using, e.g., the Newton-Raphson method [223].

In the context of this analysis, as mentioned previously, the signal final states are characterised by four hadronic jets and a large amount of missing energy. Therefore, the relevant measured quantities for the kinematic fit are the four-vectors of the four jets which are parametrised in terms of: the jet energy E, the polar (θ) and azimuthal (ϕ) angles of each jet. In order to optimise the fit, the uncertainties of these parameters were chosen to be: $\frac{\sigma_E}{E} = \frac{120\%}{\sqrt{E}}$ for the jet energy and 100 mrad for both θ and ϕ [108]. The jet uncertainties stem from the potential inaccuracies related to the quark fragmentation modelling, jet clustering, etc.

The applied kinematic constraint requires only that the masses of the two di-jet systems be equal (equal mass constraint). Since both the chargino as well as the neutralino hadronic decays are considered signal, no further mass assumptions are used as kinematic constraints.

For the Point 5 analysis, the `MarlinKinFit` package [224] was used. The presence of four hadronic jets in the final state implies that there are three possible jet pairing combinations. The kinematic fit is performed for each of the three combinations.

The quality of the fit can be evaluated by calculating the integral of the χ^2 probability distribution (prob. distrib.) [225]:

$$Prob(\chi^2; N) = \int_{\chi^2}^{\infty} P(\chi'^2; N) d\chi'^2 \text{ , where} \tag{10.13}$$

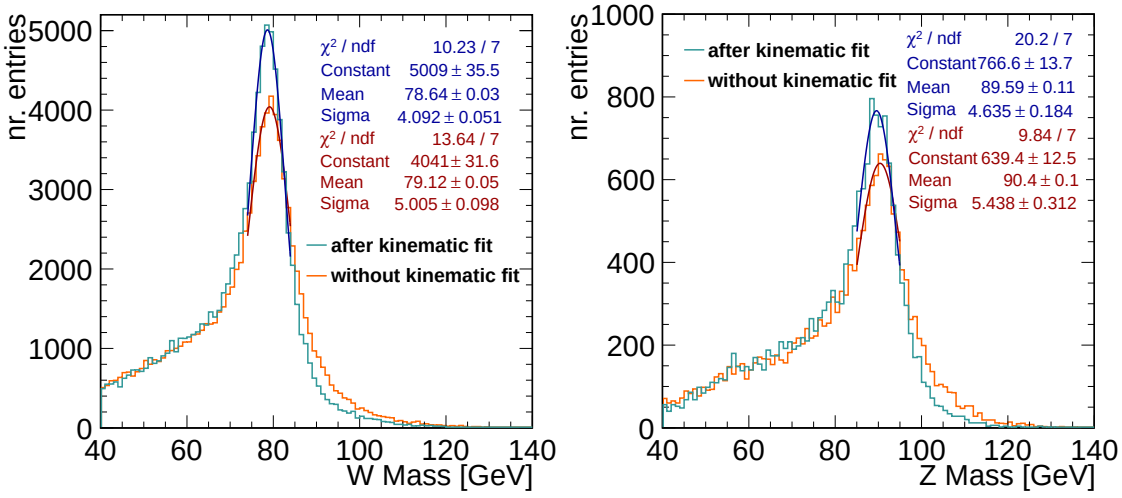
$$P(\chi^2; N) = \frac{2^{-N/2}}{\Gamma(N/2)} \chi^{N-2} e^{-\chi^2/2} \quad \text{is the } \chi^2 \text{ prob. distrib.} \tag{10.14}$$

In equation 10.14, $\Gamma(x)$ represents the typical gamma function[1] and N denotes the number of degrees of freedom. Equation 10.13 is called the *fit probability*. It expresses the probability to obtain a new χ'^2 value, in a new minimisation, which is as large or larger than the current χ^2 value. When the χ^2 value is large, the probability $Prob(\chi^2)$ is small and this typically indicates either that the errors are underestimated or that the fit hypothesis is false. On the contrary, if the

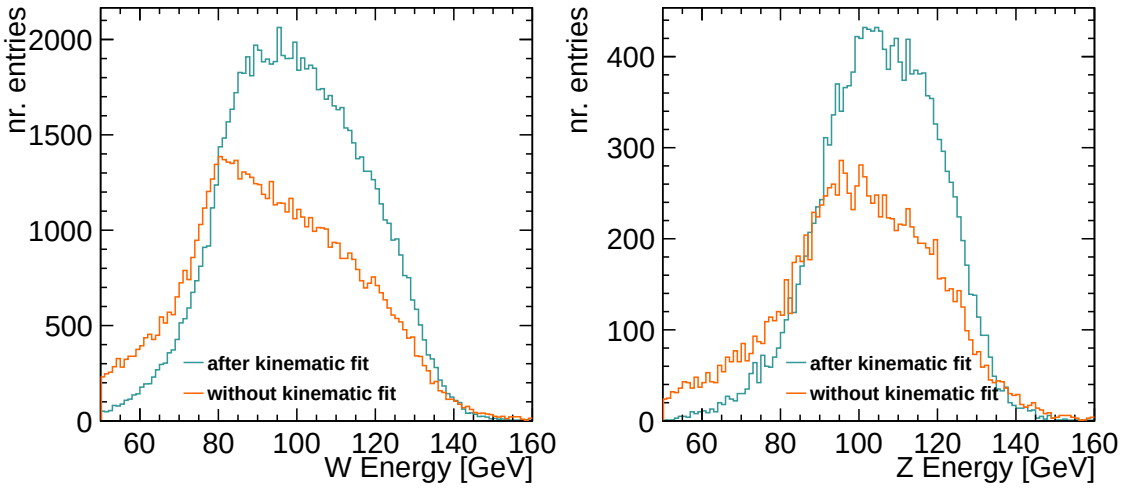
[1] The gamma function is defined as: $\Gamma(x) = (x-1)!$

χ^2 value is very small such that the probability $Prob(\chi^2)$ is very close to 1 this would indicate that the errors are overestimated. For a correct fit hypothesis and accurately estimated errors the fit probability is expected to be a uniform distribution between 0 and 1.

The fit probability is computed for each of the three possible di-jet combinations in the Point 5 analysis. The one that corresponds to the best fit probability is chosen.



(a) Chargino sample **(b)** Neutralino sample



(c) Chargino sample **(d)** Neutralino sample

Figure 10.4: Mass and energy of the reconstructed W and Z bosons, before and after the kinematic fit, as obtained from pure $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ Monte Carlo samples. The kinematic fit helps improve the dijet mass resolution with ≈ 1 GeV as well as it provides sharper box spectra for the dijet energy.

The benefits of the kinematic fit are illustrated by figure 10.4. Two pure Monte Carlo samples of $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ events, fully simulated and reconstructed with the DBD

version, were chosen to study the impact of the kinematic fitting procedure. The reconstructed boson (di-jet) masses were compared with and without applying the kinematic fit. For the case when no kinematic fit was applied the jet pairing combination which provided the smallest di-jet mass difference was chosen. In the other case, the di-jet configuration that provided the best fit probability was taken into consideration.

The distributions have been subsequently fitted with a Gaussian function from which the di-jet mass resolution was extracted. As shown in figure 10.4, the uncertainty on the W boson mass has improved from 5 GeV to 4 GeV after applying the kinematic fit. Similarly, for the Z boson mass the uncertainty has improved from 5.4 GeV to 4.6 GeV. Thus, it is clear that the kinematic fit contributes to improving the di-jet mass resolution.

The lower pannels of figure 10.4 provide a qualitative comparison of the reconstructed dijet energies with and without the kinematic fit. It is relevant to note that the positions of the edges of the boson energy spectrum are essential for the determination of the gaugino masses. Taking this into consideration, it is clearly visible that the kinematic fit contributes such that the upper and lower edges of the boson energy spectra are much better defined and thus much easier to identify and fit according to the procedures that will be described later in chapter 11. Thus, the benfits of applying a kinematic fit are observable both in terms of improving the di-jet mass resolution as well as in providing a sharper di-jet energy spectrum.

10.4 Signal Selection

In order to select the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ candidate events, several observables were chosen. Based on their distributions, a set of selection criteria (cuts) was defined. The present section describes the relevant observables and illustrates the selection cuts.

Since this analysis is based on the Point 5 study presented in [196], a major part of the selection consists in re-applying the LOI cuts on the DBD data, for legacy and comparison reasons. However, several new observables were investigated and added for the present analysis. Their aim was to improve the selection especially in view of the hadronic background originating from the low P_T $\gamma\gamma$ interactions which had not been considered in the LOI study.

The cut sequence presented below follows the exact same order in which the selection criteria are applied in the course of performing the analysis.

For each considered observable, the distributions of the signal data and the most important background contributions are illustrated using only DBD data.

The shown distributions are normalised to the assumed integrated luminosity value of 500 fb^{-1} and to the aforementioned polarisation configuration of $\text{P}_{(e^+,e^-)}$ = (+0.3, -0.8). The cuts themselves are represented by a line-crossed area, i.e., all the events found inside the marked area are discarded in that particular cut. The effects of the cuts on the event numbers and sample purities are summarised at the end of this section.

Isolated Leptons

Rejecting events that contain isolated leptons aims at significantly reducing the semileptonic and leptonic backgrounds like, e.g., $W^+W^- \to l\nu l\nu$, $W^\pm Z^0 \to l\nu ll$.

The procedure applied in order to search for and identify an isolated lepton, is described in detail in [226]. Due to the nature of the τ lepton decays, the algorithm searches only for isolated electrons and muons.

The information stored for each reconstructed particle (PFO) regarding its energy deposits in the electromagnetic calorimeter (E_{ECAL}) and/or the hadronic calorimeter (E_{HCAL}) as well as the yoke (E_{yoke}) was used. The criteria used for the electron and muon identification and separation are shown in table 10.5, where E_{total} represents the total energy deposited in all calorimeter systems while P is the reconstructed particle's measured momentum.

Lepton type	Electron	Muon
Selection	$E_{ECAL}/E_{total} > 0.9$	$E_{ECAL}/E_{total} < 0.5$
	$0.8 < E_{total}/P < 1.2$	$E_{total}/P < 0.3$
	-	$E_{yoke} > 1.2\,\text{GeV}$

Table 10.5: The lepton identification criteria.

The degree of isolation of the identified lepton was estimated by considering a cone with an opening of $\cos\theta = 0.98$ around each lepton candidate. The so called cone energy (E_{cone}) was obtained by adding up the energy of all the other PFOs that are present within the cone. If the momentum of an electron candidate fulfills the requirement that $P_{electron} > 0.25 \cdot E_{cone} \cdot 12.6\,\text{GeV}$ then it is considered an isolated electron. Similarly, an isolated muon candidate must fulfill the $P_{muon} > 0.1 \cdot E_{cone} \cdot 17.1\,\text{GeV}$ relation.

The isolated lepton identification efficiency of the described procedure was estimated to be $\approx 95\%$ for single muons and $\approx 90\%$ for single electrons, according to [227].

In the course of this analysis, the information provided by the isolated lepton identification was used in a binary way; i.e., regardless of the total number of leptons found in the event, once an isolated lepton was found that particular event was rejected.

Number of PFOs in Event

The distribution of the number of reconstructed particles present per event is illustrated in figure 10.5a for the two signal samples and the main Standard Model (SM) and SUSY backgrounds. It can be observed that the PFO multiplicities in the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ samples span the range between 30 PFOs and 150 PFOs per event. Consequently, all the events characterised by a lower or a higher reconstructed particle multiplicity, respectively, are discarded.

Number of Tracks with P_T smaller than 1 GeV

To reduce the impact of the low transverse momentum $\gamma\gamma \to hadrons$ back-

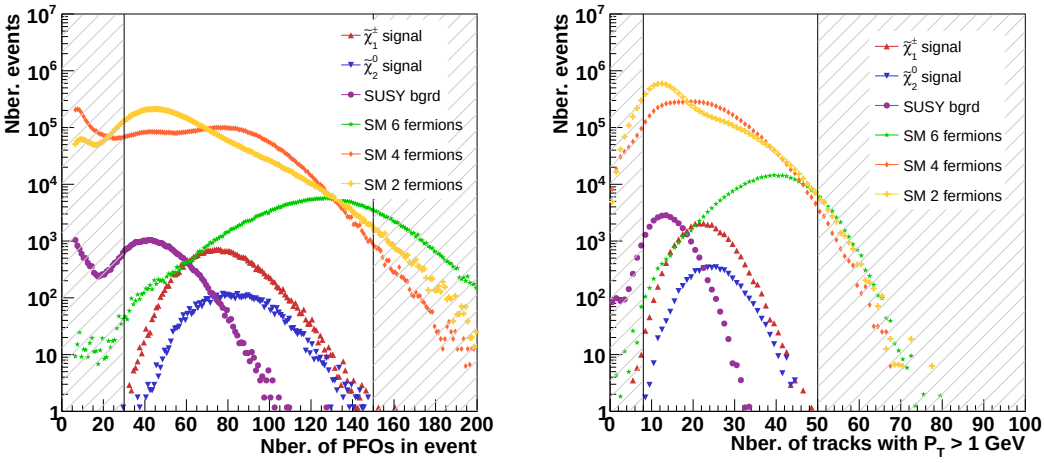


Figure 10.5: Selection cuts concerning the PFO multiplicity and the number of tracks with low P_T per event.

ground, described in section 3.3.2, a new selection criterion was considered, namely, the number of tracks reconstructed in an event that have a measured P_T lower than 1 GeV.

The distributions associated to this observable are shown in figure 10.5b. The left cut area removes those events with very few low P_T tracks, i.e. less than 8 such tracks per event. Conversely, the right hand rejection area cuts away the events which have more than 50 tracks. Apart from contributing to the $\gamma\gamma \to hadrons$ background rejection, this selection step also removes a significant amount of the Standard Model 2 and 4 fermion leptonic events.

Thrust Value

The following observable is an event shape variable that provides information on the "jet-likeness" of the event [228]. For any final state of an event, the *thrust* is defined as the normalised sum of the momenta of all reconstructed particles projected onto the event's thrust axis (n_T) which is the direction that maximises the following quantity:

$$T = \max_{|n|=1} \frac{\sum_i |\vec{p}_i \cdot \vec{n}_i|}{\sum_i |\vec{p}_i|} = \frac{\sum_i |\vec{p}_i \cdot \vec{n}_T|}{\sum_i |\vec{p}_i|} \tag{10.15}$$

The thrust observable has a value of T=1 for linear shaped events, for example, as in the case of two back-to-back hadronic jets in the final state ($Z^0 \to q\bar{q}$). For fully spherical events, in which the observed particles are symmetrically distributed, the thrust value is minimum and equal to T=0.5 [229].

The distributions of the thrust variable are shown in figure 10.6a. It can be observed that the Standard Model 2 and 4 fermions distributions peak around thrust values close to 1. The SUSY signal distributions (the red and dark blue curves) are more central and uniformly spread. Consequently, all events with a thrust value higher than 0.95 are rejected.

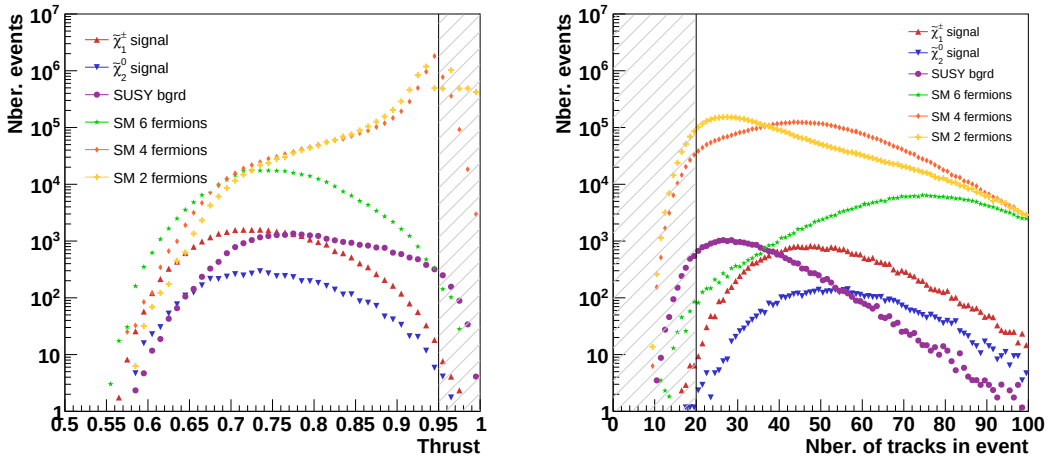


Figure 10.6: Selection criteria involving the thrust event shape variable and the track multiplicity per event.

The selection criterion based on the thrust observable was the last addition to the original LOI selection criteria [196]. The original cuts are summarised in the following.

Number of Tracks in Event

In order to further reduce the amount of SM leptonic and semileptonic background events, the number of recontructed tracks found in an event was chosen as a relevant observable. The distributions showing the track multiplicity per event are presented in figure 10.6b. It can be seen that the signal events contain a relatively higher number of tracks than the SM background samples. Therefore, in order to reject this background, all the events for which the number of measured tracks is below 20 are discarded.

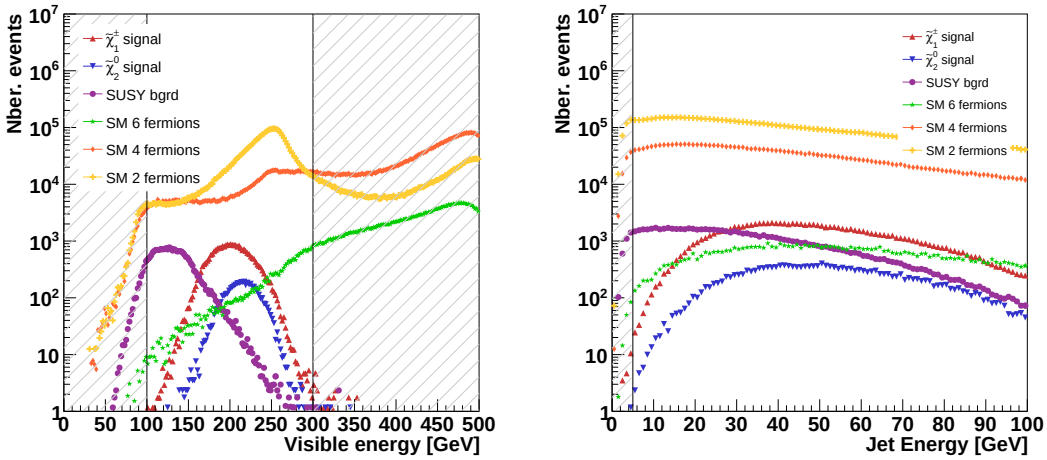
Visible Energy

Due to the presence of two invisible lightest supersymmetric particles ($\tilde{\chi}_1^0$) in the final state, the signal events are characterised by a large amount of missing energy. Therefore, the amount of reconstructed energy in the event is a very relevant selection criterion. It is defined as the sum of all the individual PFO energies.

The visible energy distributions are presented in figure 10.7a. It can be observed that the typical visible energy in a signal event can be found in the range between 100-300 GeV. Consequently, all events with a lower or higher amount of visible energy are discarded.

Jet Energy

The reconstructed jet energy distributions are shown in figure 10.7b. In order to ensure a good quality of the jet reconstruction, each of the four jets in the event should have a reconstructed energy higher than 5 GeV.

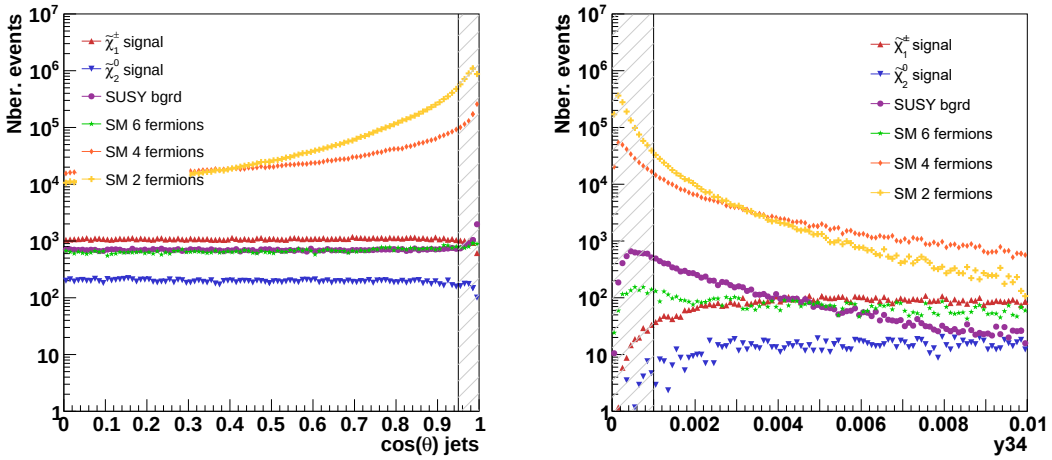


(a) (b)

Figure 10.7: Event selection criteria based on the amount of visible energy reconstructed in the event and the measured jet energy.

Jet $\cos\theta$

Another jet reconstruction quality requirement concerns the jet polar angle, measured with respect to the jet axis. The distribution of the $\cos\theta_{jet}$ observable is presented in figure 10.8a. The events containing one or more very forward jets, i.e., with $\cos\theta_{jet} > 0.99$ are rejected.



(a) (b)

Figure 10.8: Event selection cuts concerning the jet polar angle and the y_{34} parameter of the Durham jet algorithm .

Jet Clustering Parameter y_{34}

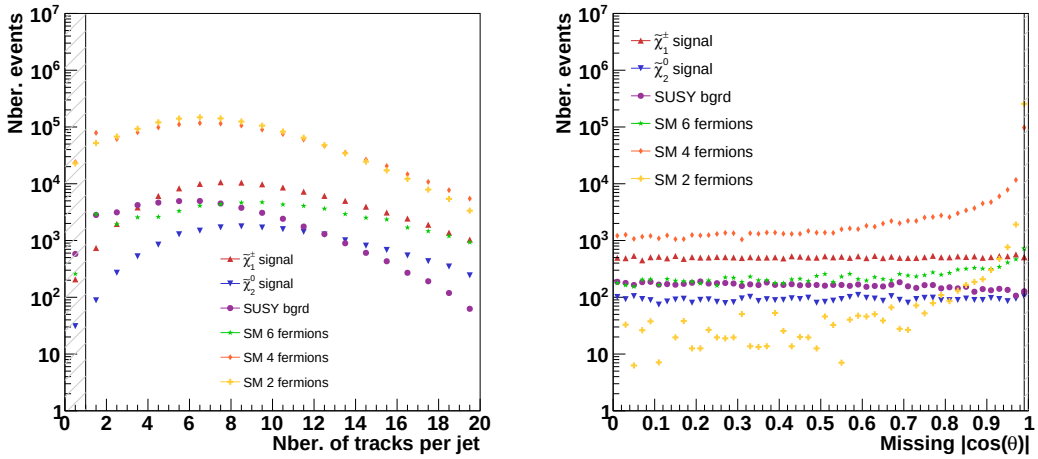
The next selection criterion involves the y_{34} parameter of the Durham jet clustering algorithm [102]. It is the distance parameter marking the transition from a four-jet to a three-jet configuration in the event.

The y_{34} distributions can be seen in figure 10.8b. The SM background events with four and especially two fermions in the final states peak towards very low

values of y_{34}. All the events for which the y_{34} value is lower than 0.001 were discarded, as marked in figure 10.8.

Number of Tracks in Jet

Another observable employed to reduce the leptonic SM background is the number of reconstructed tracks that are associated to each one of the four jets present in the event. Thus, for this selection step, the requirement is that the selected events must have at least one track in each of the four jets. The distribution of the number of tracks found in each jet and the rejection region are shown in figure 10.9a.



(a) (b)

Figure 10.9: Event selection cuts based on the number of tracks found per jet and the $\cos(\theta)$ of the missing momentum.

Missing Momentum $\cos\theta$ (1)

After grouping all the reconstructed particles in an event into four jets, each jet can be treated as a four-momentmum object itself. The jet three momentum components (p_x, p_y, p_x) as well as the energy E_j are the final result of the combination scheme employed in the jet clustering algorithm.

The missing four-momentum is defined as follows:

$$P_{miss} = \begin{pmatrix} p_{x_{miss}} \\ p_{y_{miss}} \\ p_{z_{miss}} \\ E_{miss} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ E_{cm} \end{pmatrix} - \begin{pmatrix} p_{x_1} \\ p_{y_1} \\ p_{z_1} \\ E_1 \end{pmatrix} - \begin{pmatrix} p_{x_2} \\ p_{y_2} \\ p_{z_2} \\ E_2 \end{pmatrix} - \begin{pmatrix} p_{x_3} \\ p_{y_3} \\ p_{z_3} \\ E_3 \end{pmatrix} - \begin{pmatrix} p_{x_4} \\ p_{y_4} \\ p_{z_4} \\ E_4 \end{pmatrix} \quad (10.16)$$

where E_{cm} is the ILC centre-of-mass energy, i.e., in this case, 500 GeV, and p_{x_i}, p_{y_i}, p_{z_i} and E_i with $i \in \{1 \cdots 4\}$ are the momentum and energy components, respectively, of the four jets in the event.

Thus, the $cos\theta$ of the missing momentum is:

$$\cos\theta_{miss} = \frac{p_{z_{miss}}}{\sqrt{p^2_{x_{miss}} + p^2_{y_{miss}} + p_{z^2_{miss}}}} \tag{10.17}$$

For this cut, the absolute value of the $cos\theta_{miss}$ was taken into account. Figure 10.9b, shows the distributions of this variable for all the relevant Monte Carlo samples.

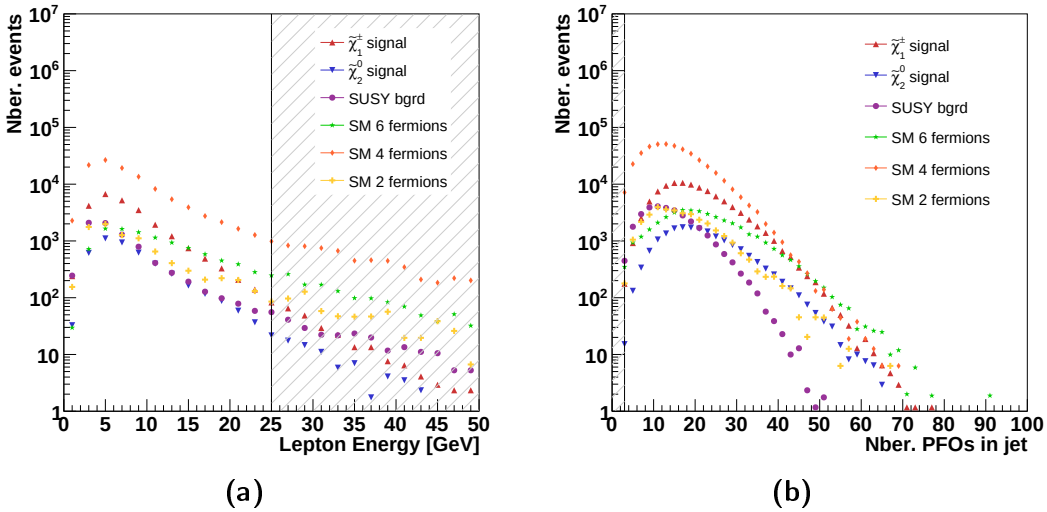
The main purpose of the $cos\theta_{miss}$ cut is to reject coplanar events like, e.g., WW with initial state radiation photons. Furthermore, as can be seen from figure 10.9b, the Standard Model background is forward peaked while the signal distributions (marked in red and dark blue) are rather uniform. All events with $cos\theta_{miss}$ larger than 0.99 are rejected.

Lepton Energy

In order to identify the non-isolated leptons (i.e., electrons and muons) present in an event the following identification procedure was employed. Any `PFO` characterised by electromagnetic (ECAL) and hadronic (HCAL) calorimeter energy deposits larger than zero and for which the relation $E_{ECAL}/(E_{ECAL}+E_{HCAL}) > 0.9$ is true is considered an electron. Conversely, if $E_{ECAL} + E_{HCAL}/E_{PFO} < 0.5$, where E_{PFO} is the total measured energy of the reconstructed particle, then the `PFO` is identified as a muon.

After the lepton identification step, the relevant observable is the energy of the highest energetic lepton in the event, E_{lepton}. The corresponding distributions are presented in figure 10.10a.

In order to further reduce the leptonic and semi-leptonic samples, all the events for which E_{lepton} is higher than 25 GeV are rejected.



(a) (b)

Figure 10.10: Event selection criteria based on the largest amount of energy carried by a lepton in the event and the number of reconstructed particles contained in a jet.

Number of PFOs per Jet

In order to further reduce the number of events containing τ decays, another variable was taken into consideration: the number of reconstructed particles associated to a jet. The distributions of this observable for the most relevant Monte Carlo samples can be seen in figure 10.10b.

Thus, aiming to reject events containing jets with low reconstructed particle multiplicities, a signal candidate must fulfill the requirement that each jet contains at least three `PFOs`. The rejection region is also marked in figure 10.10b.

Missing Momentum $\cos\theta$ (2)

In order to improve the signal selection in view of the cross-section measurements which require a higher purity of the signal samples, the $cos\theta_{miss}$ cut was applied once again with more stringent requirements.

The relevant distributions are shown in figure 10.11a. In order to eliminate a larger fraction of the SM background which typically peaks in the forward region, all the events with a $cos\theta_{miss}$ value higher than 0.8 were rejected.

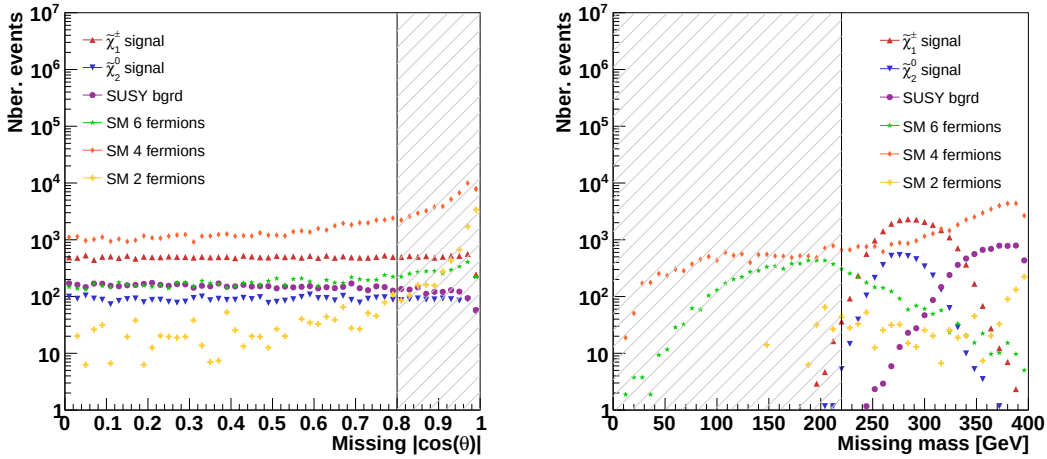


Figure 10.11: Event selection criteria based on the $\cos(\theta)$ of the missing momentum and the missing mass.

Missing Mass

The missing four-momentum was introduced and described earlier for defining the previous $cos\theta_{miss}$ cut. Using the same notation, the missing mass can then be defined as $M_{miss} = \sqrt{E_{miss} - (p^2_{x_{miss}} + p^2_{y_{miss}} + p^2_{z_{miss}})}$.

The relevant distributions of this variable can be seen in figure 10.11b. It can be observed that the SM four and six fermions events (especially semi-leptonic $t\bar{t}$) contribute most towards the lower values of the missing mass. Therefore, in order to reject these decays, all the events for which the calculated missing mass is lower than 220 GeV are rejected. The cut area can be seen in figure 10.11b.

Convergence of Kinematic Fit

In order to ensure a certain data quality such that the events can be passed on for further analysis, the kinematic fit (section 10.3) should converge for at least one jet pairing out of the possible three combinations.

Selection Results

There are two levels of the event selection employed in the Point 5 analysis. The first one, "*Level 1*", comprises the first twelve cuts. The events considered in the gaugino mass determination are required to pass (only) this level. The second selection stage, "*Level 2*", encompasses all sixteen cuts. The events selected for the cross-section measurements are required to pass the second level.

The impact of the individual selection cuts on the number of events corresponding to any of the two signal or four background samples is illustrated in table 10.6.

The results of the event selection were evaluated in terms of the efficiency, ϵ, and the percentage of events remaining in the sample after the cut. For any set of Monte Carlo data, s, and any cut, c, the efficiency is defined as:

$$\epsilon = \frac{\text{№ events } s \text{ that survive cut } c}{\text{№ events } s \text{ before cut } c} \tag{10.18}$$

The obtained efficiency and purity numbers are presented (only for the DBD sample) in table 10.7.

		Cut	$\tilde{\chi}_1^{\pm}$ hadronic (signal)	$\tilde{\chi}_2^0$ hadronic (signal)	SUSY background	2 fermions (SM)	4 fermions (SM)	6 fermions (SM)
Level 2	Level 1	No cut	27410	4898	71431	11511230	16223452	588965
		No isol. lept.	27265	4857	39566	10469255	9179111	372344
		$30 \leq$ № PFOs in evt. ≤ 150	27257	4854	28918	8782520	6678952	311856
		$8 \leq$ № tracks with $P_T > 1\,\text{GeV} \leq 50$	27212	4851	25516	7453522	6175814	282709
		Thrust < 0.95	27197	4845	24983	4544187	4929941	282435
		№ tracks in evt. ≥ 20	27177	4841	23039	4250959	4812898	282176
		$100\,\text{GeV} <$ Visible energy $< 300\,\text{GeV}$	27143	4831	20929	2858134	925843	17061
		Jet energy $> 5\,\text{GeV}$	27125	4830	17893	2537296	845343	16915
		$\lvert\cos(\theta_{Jet})\rvert < 0.99$	26515	4729	15964	1787067	607735	16051
		y_{34} parameter > 0.001	26357	4704	11202	330975	299664	14893
		№ tracks per jet > 1	25420	4585	8083	261459	205866	12125
		$\cos(\theta_{P_{miss}}) < 0.99$	25158	4535	8020	9171	117782	11656
		Lepton energy $< 25\,\text{GeV}$	24900	4460	7749	8432	109389	10121
		№ PFOs per jet > 3	24725	4445	7305	8253	102323	9783
		$\cos(\theta_{P_{miss}}) < 0.80$	19864	3589	6135	1100	53953	6955
		Missing mass $> 220\,\text{GeV}$	19826	3584	6134	931	41322	1764
		Kin. fit converged	19749	3565	5966	839	40260	1749

Table 10.6: List of the selection criteria used in the Point 5 study. The selection cuts are classified in two categories: "Level 1" corresponds to the selection stage required for the gaugino mass determination, while the extended Level 2 is employed for the cross-section measurements. The shown event numbers were obtained from DBD Monte Carlo data.

Cut	$\tilde{\chi}_1^{\pm}$ had.		$\tilde{\chi}_2^0$ had.		SUSY bg.		2f (SM)		4f (SM)		6f (SM)	
	ϵ [%]	%sample	ϵ [%]	% sample	ϵ [%]	% sample	ϵ [%]	% sample	ϵ [%]	% sample	ϵ [%]	% sample
No isol. lept.	99.47	0.14	99.18	0.02	55.39	0.20	90.95	52.11	56.58	45.68	63.22	1.85
№ PFOs in evt.	99.97	0.17	99.93	0.03	73.09	0.18	83.89	55.46	72.76	42.18	83.75	1.97
№ tracks P_T>1 GeV	99.83	0.19	99.95	0.03	88.24	0.18	84.87	53.36	92.47	44.21	90.65	2.02
Thrust	99.95	0.28	99.87	0.05	97.91	0.25	60.97	46.31	79.83	50.24	99.90	2.88
№ tracks in evt.	99.92	0.29	99.93	0.05	92.22	0.25	93.55	45.22	97.63	51.20	99.91	3.00
Visible energy	99.88	0.70	99.79	0.13	90.84	0.54	67.24	74.16	19.24	24.02	6.05	0.44
Jet energy	99.93	0.79	99.96	0.14	85.50	0.52	88.77	73.56	91.31	24.51	99.15	0.49
$\lvert\cos(\theta_{Jet})\rvert$	97.75	1.08	97.92	0.19	89.22	0.65	70.43	72.70	71.89	24.72	94.89	0.65
y_{34} parameter	99.40	3.83	99.47	0.68	70.17	1.63	18.52	48.12	49.31	43.57	92.79	2.17
№ tracks per jet	96.45	4.91	97.48	0.89	72.16	1.56	79.00	50.52	68.70	39.78	81.41	2.34
$\cos(\theta_{P_{miss}})$	98.97	14.27	98.92	2.57	99.22	4.55	3.51	5.20	57.21	66.80	96.13	6.61
Lepton energy	98.97	15.09	98.34	2.70	96.62	4.69	91.94	5.11	92.87	66.28	86.83	6.13
№ PFOs per jet	99.29	15.76	99.66	2.83	94.27	4.66	97.88	5.26	93.54	65.24	96.66	6.24
$\cos(\theta_{P_{miss}}) < 0.80$	80.34	21.69	80.76	3.92	83.98	6.70	13.33	1.20	52.73	58.90	71.10	7.59
Missing mass	99.81	26.95	99.85	4.87	99.98	8.34	84.63	1.27	76.59	56.17	25.36	2.40
Kin. fit converged	99.61	27.38	99.48	4.94	97.27	8.27	90.17	1.16	97.43	55.82	99.15	2.42

Table 10.7: Performance of the selection cuts in terms of the efficiency and number of events remaining in the *total* sample.

CHAPTER 11

Mass Measurements

The present chapter describes two analysis methods that were employed in the determination of the gaugino ($\tilde{\chi}_1^{\pm}$, $\tilde{\chi}_2^0$, $\tilde{\chi}_1^0$) masses.

The first method was proposed in the original Point 5 study published in the ILD *Letter of Intent* document [109, 196] and was applied on the LOI data set. For comparison purposes, it was also carried out using the DBD data set and the results are described in the following sections. A second, new and more robust approach that emerged in the course of this thesis is also presented in detail.

In the following sections, both analysis procedures are illustrated using the DBD data set. For this purpose, the data underwent the preparation stage described in chapter 10. Only those events that survived the *Level 1* stage of the selection were used in the analysis.

The underlying principle of the two methods described in this chapter is that the gaugino masses can be determined using the positions of the edges exhibited by the relevant energy spectra. The two approaches essentially extract the input information necessary for the mass calculation from the data. The calculation itself is presented in the following section.

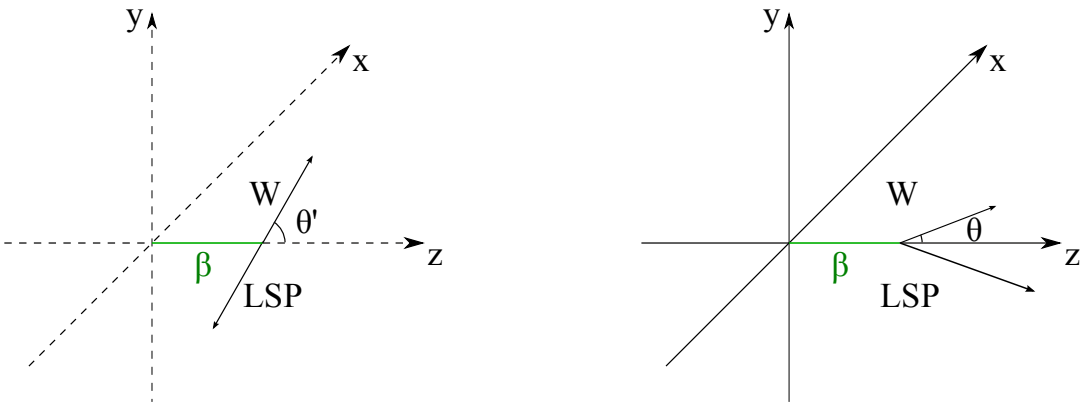
11.1 Mass Determination from Energy Spectra Edge Positions

11.1.1 The Boson Energy Spectrum

The signal processes in the Point 5 physics scenario are of the type $M_\chi \rightarrow M_{LSP} + M_V$, where M_χ represents either the $\tilde{\chi}_1^{\pm}$ or $\tilde{\chi}_2^0$, M_{LSP} denotes the lightest supersymmetric particle (LSP) $\tilde{\chi}_1^0$ and M_V stands for the corresponding gauge boson. Furthermore, the mass difference between M_χ and M_{LSP} is large enough that the produced gauge bosons are real, on-shell particles. Therefore, the starting

point of the discussion is the well known kinematics of two body decays.

For the following discussion, the standard ILD detector cartesian coordinate system was considered. The origin of the coordinate system is the interaction point while the z axis is collinear to the direction of the electron beam and the y axis lies vertically and points upwards.



(a) Gaugino rest frame **(b)** Lab frame

Figure 11.1: Sketch of gaugino decay kinematics.

The four-momenta of the particles involved in the process can be written as: $\mathbf{P}_\chi = (E_\chi, \vec{p}_\chi)$, $\mathbf{P}_V = (E_V, \vec{p}_V)$ and $\mathbf{P}_{LSP} = (E_{LSP}, \vec{p}_{LSP})$. The energy and momentum conservation clearly requires that $\mathbf{P}_\chi = \mathbf{P}_V + \mathbf{P}_{LSP}$. From this, the four-momentum of the LSP can be expressed as $\mathbf{P}_{LSP} = \mathbf{P}_\chi$ - $\mathbf{P}_V$. Now, squaring both sides of the previous expression, the following is obtained:

$$\mathbf{P}_{LSP}^2 = (\mathbf{P}_\chi - \mathbf{P}_V)^2 = \mathbf{P}_\chi^2 + \mathbf{P}_V^2 - 2\mathbf{P}_\chi\mathbf{P}_V \tag{11.1}$$

Since the norm of a four-momentum yields the invariant mass of the respective particle and after applying the multiplication rules of four-vectors, equation 11.1 can be re-written as:

$$M_{LSP}^2 = M_\chi^2 + M_V^2 - 2(E'_\chi E'_V - \vec{p}_\chi\vec{p}_V) \tag{11.2}$$

In the gaugino rest frame, schematically illustrated in figure 11.1 and denoted with an apostrophe symbol, the mother particle's momentum is zero:

$$E'_\chi = M_\chi, \quad \vec{p}_\chi = \vec{0} \;\Rightarrow \mathbf{P}'_\chi = (M_\chi, \vec{0}) \tag{11.3}$$

Solving equation 11.2 for E'_V while noting that $\vec{p}_\chi = \vec{0}$ yields:

$$E'_V = \frac{M_\chi^2 + M_V^2 - M_{LSP}^2}{2E'_\chi} \overset{(E'_\chi = M_\chi)}{=} \frac{M_\chi^2 + M_V^2 - M_{LSP}^2}{2M_\chi} \tag{11.4}$$

In the next step, the expression for the gauge boson energy (11.4) can be boosted into the laboratory system, in which the relevant observables are measured. The Lorentz boost coefficient γ is determined by: $\gamma = E_\chi/M_\chi$ where E_χ and M_χ represent the energy and mass of the decaying gaugino. This definition of

γ is valid under the assumption that typical radiation losses can be ignored such that the energy of the gaugino is equal to the beam energy. Thus, the following expression can be obtained:

$$\begin{aligned} E_V &= \gamma E'_V + \beta\gamma \vec{p}_V \\ &= \gamma E'_V + \beta\gamma |\vec{p}_V| \cos(\theta') \end{aligned} \tag{11.5}$$

where θ' is the angle formed by the gauge boson's momentum with respect to the direction of the boost, in the centre of mass frame. In figure 11.1 the boost direction is illustrated as parallel to the z-axis for simplicity.

From equation 11.5, it can be seen that the energy of the gauge boson, in the laboratory frame, depends on the value of the θ' angle:

$$\begin{aligned} &\text{If } \theta' = 0 \Rightarrow (\cos(0) = 1) \quad \Rightarrow E_V = \gamma E'_V + \beta\gamma |\vec{\mathbf{P}}'_V| \Rightarrow E_h \\ &\text{If } \theta' = \pi \Rightarrow (\cos(\pi) = -1) \Rightarrow E_V = \gamma E'_V - \beta\gamma |\vec{\mathbf{P}}'_V| \Rightarrow E_l \end{aligned} \tag{11.6}$$

Thus, in a real experiment, the measured gauge boson energies would form a distribution with a minimum edge (E_l) that corresponds to the $\theta' = \pi$ configuration while the maximum edge (E_h) corresponds to the $\theta' = 0$ case.

The shape of the energy spectrum depends on the spin of the mother particle. For instance, if the decaying particle were a sfermion, i.e., a sparticle with spin 0, the decay would be isotropic and the energy spectrum of the visible decay products would have a flat "box"-like shape.

In the case of gauginos, which have spin 1/2, the description of the shape of the energy spectrum is more involved. It depends on the reference frame in which the distribution is measured and the orientation of the decaying particle's spin in that particular frame. A detailed discussion of the spectrum shape in the context of the gaugino decays is beyond the scope of this section.

Figure 11.2 illustrates the di-jet (boson) energy distributions obtained from the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ Monte Carlo signal samples on the generator level. It can be observed that the edges are not clear-cut. The smearing effects reflect the unavoidable consequences of the beam energy spectrum, the emission of initial state radiation and the fact that the gauge bosons have a natural width, $\Gamma_{W,Z}$.

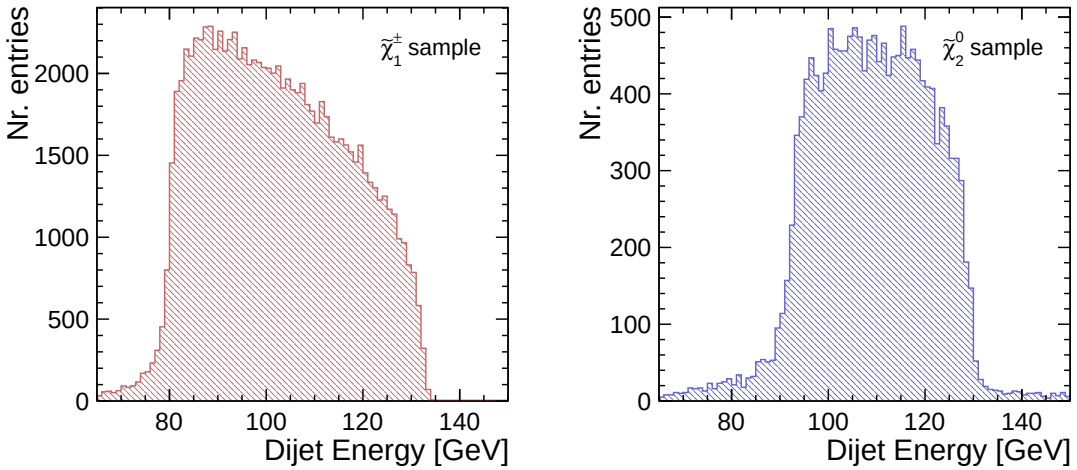
For the $\tilde{\chi}_1^{\pm}$ di-jet energy distribution, the position of the lower edge of the spectrum essentially overlaps with W boson mass. This situation arises when the momentum of the W compensates the initial boost such that the W appears to be produced at rest while the LSP carries away all the available energy.

11.1.2 Gaugino Mass Calculation

The main goal is to determine the values of the gaugino masses from the measured edge positions of the boson energy spectra. The following paragraphs present the deduction of the relevant equations.

For this purpose, two variables were introduced:

$$E_A = \frac{E_h + E_l}{2} \text{ and } E_S = \frac{E_h - E_l}{2} \tag{11.7}$$



(a) $\tilde{\chi}_1^{\pm}$ sample **(b)** $\tilde{\chi}_2^0$ sample

Figure 11.2: Boson energy spectra obtained from the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ signal samples on generator level.

where E_A represents the centre point in the allowed energy range while E_S corresponds to its width. It is useful to consider the results of the addition and subtraction of E_h and E_l:

$$E_h + E_l = \gamma E'_V + \gamma\beta\vec{p}_V + \gamma E'_V - \gamma\beta\vec{p}_V \tag{11.8}$$

$$E_h + E_l = 2\gamma E'_V \Rightarrow E_A = \gamma E'_V \tag{11.9}$$

$$E_h - E_l = \gamma E'_V + \gamma\beta\sqrt{E'^2_V - M^2_V} - \gamma E'_V + \gamma\beta\sqrt{E'^2_V - M^2_V} \tag{11.10}$$

$$E_h - E_l = 2\gamma\beta\sqrt{E'^2_V - M^2_V} \Rightarrow E_S = \gamma\beta\sqrt{E'^2_V - M^2_V} \tag{11.11}$$

In the next step, the first result of equation 11.11, namely that $E_A = \gamma E'_V$, is used to rewrite the expression for E_S:

$$\begin{aligned} E_S &= \gamma\sqrt{1 - \frac{1}{\gamma^2}}\sqrt{E'^2_V - M^2_V} \\ &= \gamma\sqrt{1 - \frac{1}{\gamma^2}}\sqrt{\left(\frac{E_A}{\gamma}\right)^2 - M^2_V} \\ &= \sqrt{1 - \frac{1}{\gamma^2}}\sqrt{E^2_A - \gamma^2 M^2_V} \end{aligned} \tag{11.12}$$

By squaring both sides of equation 11.12, a quadratic equation where γ^2 is

treated as the unkown can be obtained:

$$\begin{aligned} E_S^2 &= \left(1 - \frac{1}{\gamma^2}\right)\left(E_A^2 - \gamma^2 M_V^2\right) \\ &= E_A^2 - \gamma^2 M_V^2 - \frac{E_A^2}{\gamma^2} + M_V^2 \mid \times \gamma^2 \qquad (11.13) \\ E_S^2\gamma^2 &= E_A^2\gamma^2 - \gamma^4 M_V^2 - E_A^2 + \gamma^2 M_V^2 \\ 0 &= \gamma^4 M_V^2 + \gamma^2(E_S^2 - E_A^2 - M_V^2) + E_A^2 \qquad (11.14) \\ 0 &= \gamma^4 M_V^2 + \gamma^2(E_h E_l + M_V^2) + E_A^2 \qquad (11.15) \end{aligned}$$

The solutions for γ^2 from the quadratic equation shown above are then of the form:

$$\begin{aligned} \gamma_{1,2}^2 &= \frac{(E_h E_l + M_V^2) \pm \sqrt{(M_V^2 + E_h E_l)^2 - 4M_V^2 E_A^2}}{2M_V^2} \\ &= \frac{(E_h E_l + M_V^2) \pm \sqrt{(E_h^2 - M_V^2)(E_l^2 - M_V^2)}}{2M_V^2} \qquad (11.16) \end{aligned}$$

Since, according to equation 11.11, $\gamma E_V' = E_A$, the boson energy (in the centre of mass reference frame) can be replaced in this expression by the quantity defined in equation 11.4. Thus, the following equality can be obtained:

$$\gamma \frac{M_\chi^2 + M_V^2 - M_{LSP}^2}{2M_\chi} = \frac{E_h + El}{2} \qquad (11.17)$$

The relevant information to be extracted from the previous expression is the mass of the lightest supersymmetric particle (M_{LSP}). Therefore, the equation can be re-written such that:

$$\begin{aligned} \gamma M_\chi^2 + \gamma M_V^2 - \gamma M_{LSP}^2 &= M_\chi(E_h + E_l) \\ M_{LSP}^2 &= M_\chi^2 + M_V^2 - \frac{M_\chi}{\gamma}(E_h + E_l) \qquad (11.18) \\ M_{LSP}^2 &\overset{M_\chi = \frac{E_{beam}}{\gamma}}{=} \left(\frac{E_{beam}}{\gamma}\right)^2 + M_V^2 - \frac{E_{beam}}{\gamma^2}(E_h + E_l) \qquad (11.19) \\ M_{LSP}^2 &= M_V^2 + \frac{E_{beam}^2}{\gamma^2}\left(1 - \frac{E_h + E_l}{E_{beam}}\right) \qquad (11.20) \end{aligned}$$

The lower edge of the W energy spectrum overlaps with the W boson mass, consequently, its measured value is not used in the mass calculation.

Firstly, the value of γ is computed from equation 11.16, plugging in the two edge positions determined from the Z boson spectrum. Considering the constraint that the LSP mass is identical both for the $\tilde{\chi}_1^\pm$ as well as for the $\tilde{\chi}_2^0$ decays, the plus sign in equation 11.16 provides the unique solution for computing γ.

The $\tilde{\chi}_0^2$ mass is then calculated as $M_{\tilde{\chi}_0^2} = E_{beam}/\gamma$. The LSP mass, $M_{\tilde{\chi}_1^2}$ is obtained from equation 11.17 using the edge values determined from the Z di-jet energy distribution.

Lastly, the $\tilde{\chi}_1^\pm$ mass is determined using the previously calculated LSP mass and the value of the upper edge of the W boson energy spectrum.

11.2 Signal Separation

The events used in the gaugino mass determination are the ones that survive the Level 1 stage of the event selection presented in section 10.4. In order to separate the data sample into potential $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ candidates, a procedure based on two χ^2 calculations was employed. The important observable for the separation is the calculated di-jet (boson) mass *before* running the kinematic fit. The information obtained from the kinematic fit result is still used as only the di-jet masses which correspond to the jet pairing configurations with the best kinematic fit probability were used.

Two χ^2 variables based on the di-jet mass observables described above have been defined:

$$\begin{aligned} \chi^2_W &= \frac{(M_{dijet1} - M_W)^2 + (M_{dijet2} - M_W)^2}{\sigma^2} \\ \chi^2_Z &= \frac{(M_{dijet1} - M_Z)^2 + (M_{dijet2} - M_Z)^2}{\sigma^2} \end{aligned} \tag{11.21}$$

where χ^2_W and χ^2_Z are the χ^2 variables calculated with respect to the W (M_W) and Z (M_Z) boson masses and M_{dijet1} and M_{dijet2} represent the di-jet masses. The σ is an estimate of the detector resolution. For consistency, the same σ value was used as in the original study performed for and published in the ILD *Letter of intent* document [196]. Consequently, the value was fixed to $\sigma = 5\,\text{GeV}$.

For each event, the two χ^2 values are computed. Figure 11.3 shows the distribution of the computed χ^2_W values plotted with respect to the χ^2_Z ones.

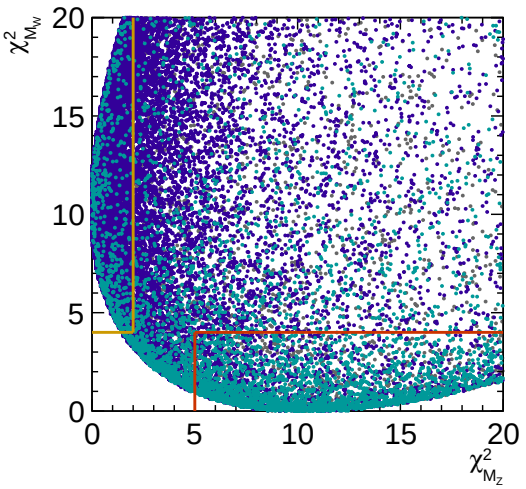


Figure 11.3: Distribution of the two calculated χ^2 variables plotted with respect to each other. Based on Monte Carlo truth information, the cyan-coloured points represent $\tilde{\chi}_1^\pm$ events while the dark blue ones represent $\tilde{\chi}_2^0$. The yellow area corresponds to the $\tilde{\chi}_2^0$ selection while the red box represents the $\tilde{\chi}_1^\pm$ selection.

The colour code is based on Monte Carlo truth information: the cyan dots represent (true) chargino signal events while the blue points correspond to (true)

neutralino events. Their distribution, shown in figure 11.3, informs the definition of the selection regions. Consequently, if the χ^2_W value is smaller than 4, then the event is considered a chargino candidate (red box). If χ^2_W is larger than 4 and χ^2_Z is simultaneously smaller than 2 the event is classified as a neutralino candidate (yellow box). Events with significantly larger χ^2 values are not considered in the analysis.

In order to evaluate the performance of the signal sample separation procedure, the efficiency (ϵ) and purity (π) of the final samples was computed according to the definition shown in section 10.4. The purity was computed both with respect to the entire data sample, i.e., also containing the Standard Model background, and only with respect to the SUSY data sample. The results are summarised in table 11.1.

Observable	$\tilde{\chi}_1^{\pm}$ sample	$\tilde{\chi}_2^0$ sample
Efficiency	53%	30%
Purity (total)	63%	39%
Purity (SUSY)	94%	62%

Table 11.1: The efficiency and purity of the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ selection.

It must be noted that, as can be seen both from figure 11.3 and table 11.1, a large part of the events is not included in the separated signal regions. The size of the data samples after the separation procedure described in this section is $N_{\tilde{\chi}_1^{\pm}} = 10519$ events for the charginos and $N_{\tilde{\chi}_2^0} = 1081$. Despite the decrease in the data sample, the purity levels achieved after applying the signal separation procedure have improved considerably: by almost a factor of three for the chargino case and approximately by a factor of eight in the neutralino selection.

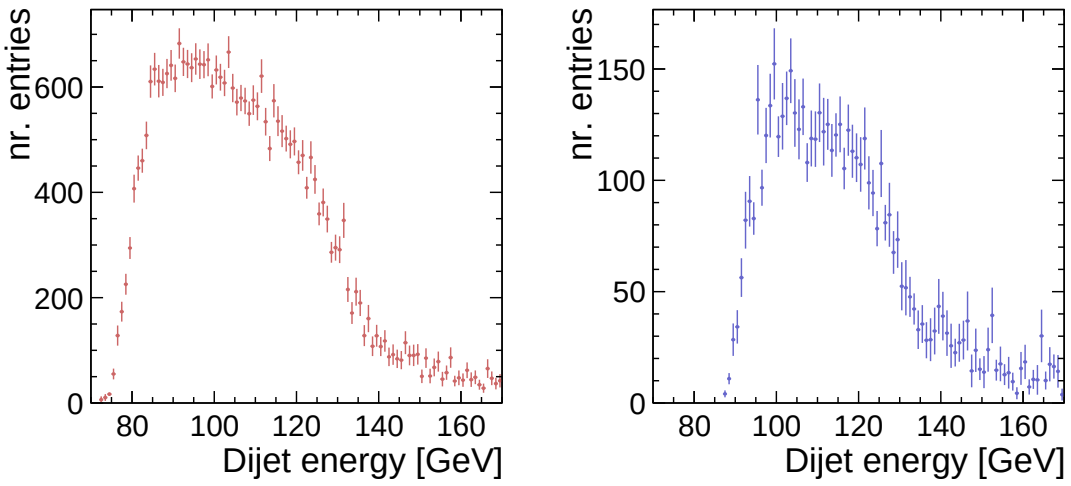
From these separated data samples, the edge position information from the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ energy spectra can be extracted. The following sections describe in detail two different procedures that were employed for this purpose.

11.3 Fitting the Edge Positions

In the original Point 5 study [196], the positions of the edges exhibited by the di-jet energy spectra were determined by fitting the aforementioned spectra with a function in which the edge positions were free parameters. For comparison and legacy reasons, the same fitting procedure was applied in the present analysis as well. The remainder of this section describes the method itself and presents the results obtained from applying it to the DBD Monte Carlo samples.

The relevant observable for the fitting method is the gauge boson (di-jet) energy *after* running the kinematic fit. Figure 11.4 shows the di-jet energy distributions for the chargino and the neutralino candidate events, after the event selection (Level 1) and the χ^2-based signal separation. They are the distrbutions on which the fit should be performed.

The histograms shown in figure 11.4 comprise $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ signal events, but also contributions from other SUSY and Standard Model backgrounds that survive the



(a) $\tilde{\chi}_1^{\pm}$ candidates sample **(b)** $\tilde{\chi}_2^0$ candidates sample

Figure 11.4: Di-jet energy spectra after performing the signal separation presented in section 11.2.

event selection and separation. The presence of the SM background is particularly visible in the region of the upper edge ($\geq$135 GeV) of the spectra.

When defining the fit function, the assumption was made that the Standard Model (SM) background is well known and its contributions to the distributions in figure 11.4 can be accurately accounted for by a component of the total fit function. The function used to describe the SM contributions is presented in the following.

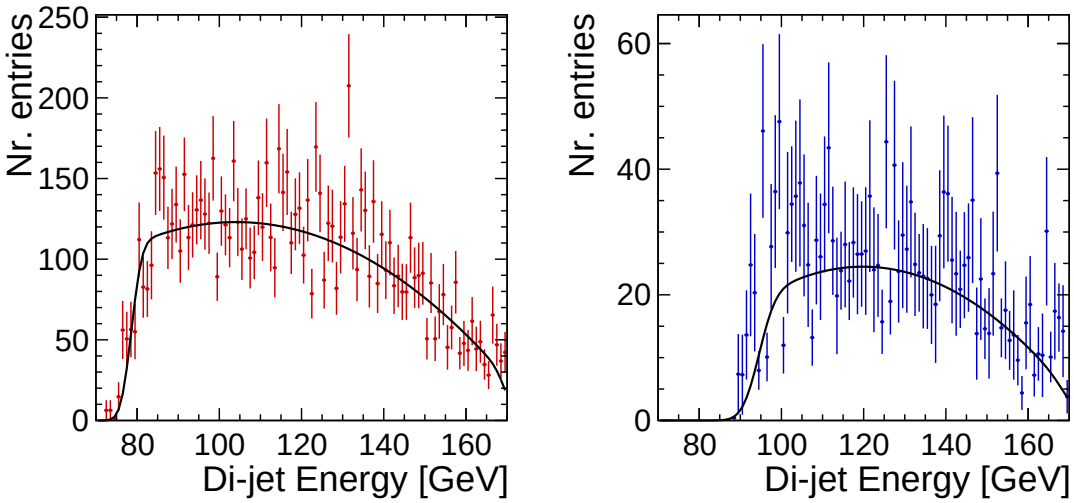
The Fit Function for the Standard Model Background

The di-jet energy distributions obtained from the Standard Model events still present in the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ candidates samples are shown as histograms in 11.5. The large error bars illustrate that the number of surviving SM events is rather small.

Since there is no specific underlying physics model that can be used for fitting the Standard Model background distributions, several polynomial functions of different orders have been tried in order to describe the shape of the descending plateau. A second order polynomial was finally chosen as providing the best fit performance. The detector resolution and the natural gauge boson width are taken into account by convoluting the polynomial with a Voigt function. Thus the fit function for the SM background was defined as:

$$f_{SM} = \int_{t_0}^{\infty} (a_{SM} \cdot t^2 + b_{SM} \cdot t + c_{SM}) Voigt(t - x, \sigma, \Gamma) dt \tag{11.22}$$

where a_{SM}, b_{SM} and c_{SM} are the polynomial coefficients, σ denotes the detector resolution and Γ represents the repective gauge boson width.



(a) Chargino selection **(b)** Neutralino selection

Figure 11.5: Di-jet energy distributions obtained from Standard Model events that contaminate the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ event selection. The black lines illustrate the result of fitting the two distributions with the function given in equation 11.22.

The Fit Function for the Complete Sample

The fitting function for the *signal* sample is defined analogously, with a second order polynomial describing the top part of the box spectra. Both the natural width of the gauge bosons and the uncertainties related to measuring the jet energies have a smearing effect on the edge position. They are taken into account, as previously, via a Voigt function.

The *total* fit function is defined as a linear combination of the functions describing the Standard Model background (equation 11.22) and the signal contribution:

$$f_{total} = f_{SM} + \int_{t_0}^{t_1} (a_{sig} \cdot t^2 + b_{sig} \cdot t + c_{sig}) Voigt(t - x, \sigma, \Gamma) dt \tag{11.23}$$

The convolution limits (t_0 and t_1) represent the edge positions and they are free parameters of the function.

Since the detector resolution (σ) can vary for different jet energies, the assumption was made that the dependence is linear. Thus, instead of employing a fixed value for σ for the whole range of the fitting function, the resolution values were allowed to vary especially around the edge positions. The expression that was used to define σ is:

$$\sigma = \sigma_0 + \frac{(\sigma_1 - \sigma_0)(t - 80)}{40} \tag{11.24}$$

where σ_0 and σ_1 are additional free parameters of the fitting function and t represents the threshold variable.

The parameter values of the f_{SM} component from the total fit function must be first determined and then fixed in equation 11.23. Ideally, the function f_{SM} (equation 11.22) would be fit to a sample of simulated Standard Model events

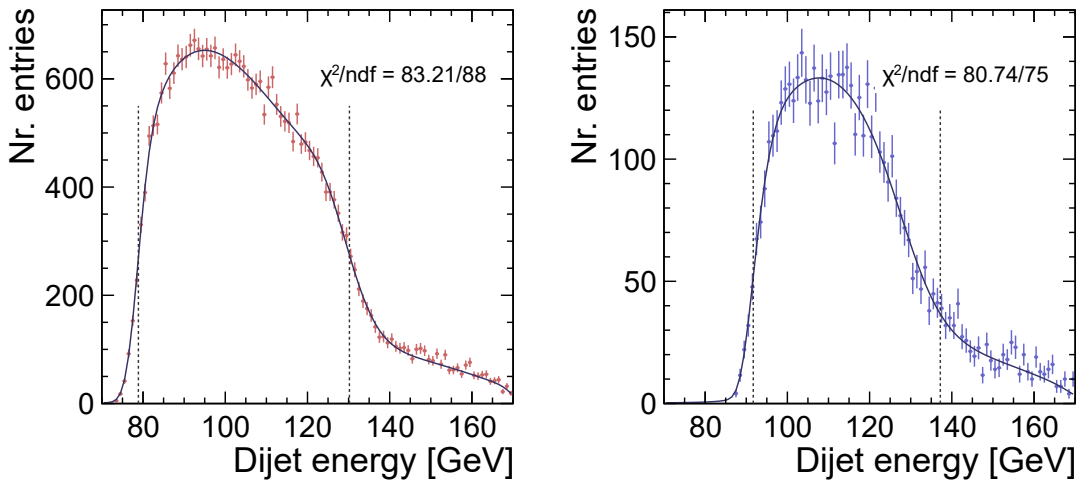
that is completely statistically independent from the events comprising the distributions in figure 11.5. Due to the limited amount of Monte Carlo data, this is not possible. Consequently, a different strategy was used.

Firstly, the Standard Model fit function (equation 11.22) was applied to the di-jet energy distributions shown in figure 11.5. The fit results are illustrated by the continuous black lines shown in the same figure. For the $\tilde{\chi}_1^{\pm}$ case the ratio between the fit χ^2 and the numbers of degree of freedom is $\chi^2/ndf = 117.6/92 = 1.3$ while in the $\tilde{\chi}_2^0$ case the value is $\chi^2/ndf = 88.06/76 = 1.2$. The function parameter values obtained from the fit were then fixed and implemented in equation 11.23.

The total fit function (f_{total}) could now, in principle, be applied to the distributions in figure 11.4. However, they contain the Standard Model events already used in the f_{SM} fit. In order to avoid introducing any unwanted bias, two new distributions of Standard Model events (one for each gaugino candidate sample) were randomly generated. This was achieved by sampling the function 11.22, in which the previously determined parameter values were fixed. The newly obtained SM di-jet distributions are then added to the SUSY signal and SUSY background events from figure 11.4, essentially replacing the original SM contributions. Finally, the total fit function can then be applied to the resulting energy spectra. It must be noted that the same method of randomly producing new Standard Model di-jet energy distributions was also employed in the initial (LOI) implementation [196] of the analysis in order to handle the limited amount of Monte Carlo data.

The Fit Results

The di-jet energy spectra containing the randomly generated Standard Model contributions and the fit results are presented in figure 11.6. The edge positions determined from the fit are illustrated by the dashed vertical lines.



(a) $\tilde{\chi}_1^{\pm}$ selection **(b)** $\tilde{\chi}_2^0$ selection

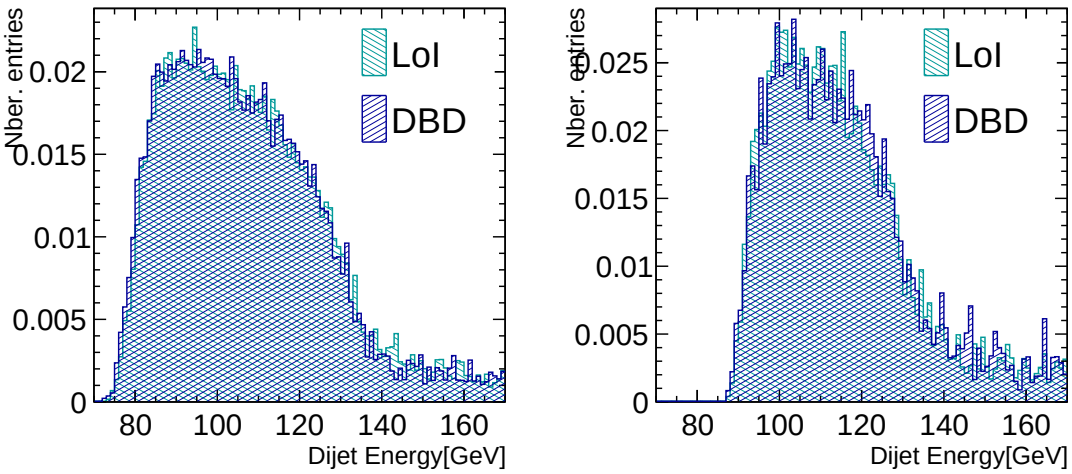
Figure 11.6: The newly obtained di-jet energy spectra. The result of fitting the distributions with the total fit function (equation 11.23) is shown as the continuous black line. The dashed vertical lines indicate the edge positions obtained from the fit.

It can be observed that the fit appears to perform well in describing the distributions, as indicated by the small value of the χ^2/ndf ratio: $\chi^2/ndf < 1.1$ for both gaugino selections.

The obtained edge values are summarised in table 11.2. The numbers achieved in the original LOI study [196] ($\text{LOI}_{\text{original}}$) are also shown for comparison.

Edge	Fit result: DBD (GeV)	Fit result: $\text{LOI}_{\text{original}}$ (GeV)
$\tilde{\chi}_1^{\pm}$ low	78.9 ± 0.3	79.9 ± 0.2
$\tilde{\chi}_1^{\pm}$ high	130.2 ± 0.7	131.5 ± 0.7
$\tilde{\chi}_2^0$ low	91.7 ± 0.4	92.3 ± 0.4
$\tilde{\chi}_2^0$ high	137.2 ± 5.4	127.7 ± 0.8

Table 11.2: Edge positions determined from fitting the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ di-jet energy spectra with the total fit function given in equation 11.23. Both calculated estimates and the values obtained in the previous LOI study [196] are shown for comparison.



(a) $\tilde{\chi}_1^{\pm}$ selection **(b)** $\tilde{\chi}_2^0$ selection

Figure 11.7: Comparison of the di-jet energy spectra obtained from the DBD and $\text{LOI}_{\text{original}}$ data sets. The distributions were normalised to an integral of one.

The comparison of the DBD and $\text{LOI}_{\text{original}}$ results[1] shown in table 11.2 reveals that while the values for the $\tilde{\chi}_1^{\pm}$ upper edge and the $\tilde{\chi}_2^0$ lower edge are compatible within their uncertainties, the other two edge values are discrepant. The largest difference, of about 10 GeV, concerns the $\tilde{\chi}_2^0$ high edge.

In view of this large discrepancy, the DBD and $\text{LOI}_{\text{original}}$ di-jet energy distributions were compared, on first principles, as illustrated in figure 11.7. It can be seen that for both the $\tilde{\chi}_1^{\pm}$ and the $\tilde{\chi}_2^0$ selections the distributions are very similar.

[1] The uncertainties quoted for the LOI results were determined from a toy Monte Carlo study in which 100 new di-jet energy distributions were randomly generated and fitted. In contrast, for the DBD results shown in table 11.2 the fit was performed only once.

The ≈10 GeV difference indicated by the fit results in the case of the $\tilde{\chi}_2^0$ high edge does not reflect a real shift of that magnitude between the DBD and the $\text{LOI}_{\text{original}}$ di-jet energy spectra.

Performing the fit on the $\tilde{\chi}_2^0$ di-jet energy distribution is more problematic due to its lower statistics. Furthermore, the upper edge in particular is typically obscured by the rather large presence of the Standard Model background. In the DBD case, the fit wrongly identified the inflection point where the SM contribution becomes dominant as the edge position (figure 11.6). This must be contrasted with the very good value of the χ^2/ndf ratio presented above. The issues in identifying the $\tilde{\chi}_2^0$ upper edge are also reflected in the significantly (≈7 times) larger uncertainty on its position extracted from the DBD fit. The failed edge identification led to the large discrepancy observed with respect to the $\text{LOI}_{\text{original}}$ value. This provided a first indication of the less than optimal performance of the fitting procedure.

Fit Consistency Test

In order to evaluate the stability of the previously discussed fitting method for extracting the edge positions from the di-jet energy spectra, a toy Monte Carlo study was performed in this thesis. For this purpose, 10^4 new di-jet energy distributions (for each of the two gaugino candidates' samples) were randomly produced based on the ones shown in figure 11.6 as follows.

Firstly, the number of entries that the new histogram was required to have was determined. This was achieved by sampling a Poisson distribution centred around the number of entries of the original histogram. Secondly, the number of entries in each bin of the new energy spectrum was obtained by sampling a Poisson distribution, this time centred around the number of entries in the corresponding bin of the original distribution. This was performed until the requirement concerning the total number of entries in the new histogram was fulfilled.

This process was repeated 10^4 times, thus producing 10^4 new di-jet energy spectra for each gaugino selection. The fitting function was then applied on each individual, newly produced, energy spectrum.

From 10^4 performed fits, 8808 have converged for the chargino case and 8282 for the neutralino case. The results of the toy Monte Carlo study are illustrated in figure 11.8. Only the converged fits are shown.

In figure 11.8, the χ^2/ndf ratio, as a measure of the fit performance, is plotted with respect to the determined edge position. The information concerning the number of fits associated with a specific χ^2/ndf - edge position configuration can be read from the colour code.

The instability of the fitting method is first suggested by the wide range of obtained edge values. For instance, in the case of the $\tilde{\chi}_2^0$ lower edge, shown in pannel (b), the obtained edge values vary over a range of ≈10 GeV despite the apparently good quality of the fit ($\chi^2/ndf < 1.5$). This effect is significantly larger than any expected shift in the edge position that could occur from the random generation of the distribution.

A clear indication of the method's instability is provided by the significantly

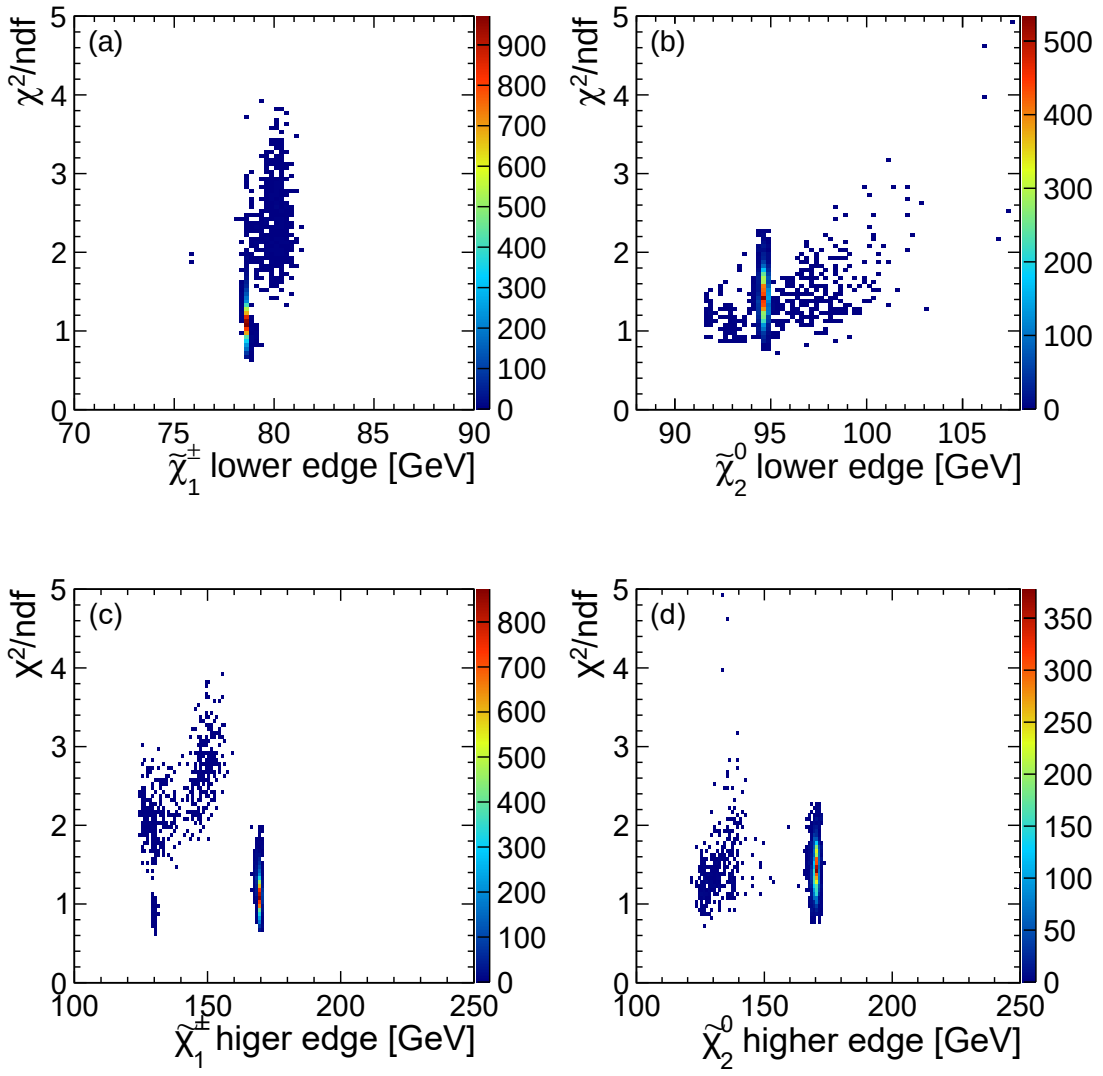


Figure 11.8: The results of the toy Monte Carlo study carried out to evaluate the consistency of the fitting method for measuring the di-jet energy spectrum edge positions. The number of converged fits is illustrated by the colour code.

worse outcome of fitting the more difficult upper edges (pannels (c) and (d)). In this case, the majority of fits falsely determined that the higher edge positions coincide with the histrograms' end points at ≈170 GeV. This is true even for high quality fits with $\chi^2/ndf < 1.5$.

Thus, as a result of the toy Monte Carlo study it could be assessed that the fitting method is highly sensitive to small fluctuations in the di-jet energy spectra. It was due to this instability that the previously shown DBD and $\text{LOI}_{\text{original}}$ results concerning the $\tilde{\chi}_2^0$ upper edge (table 11.2) were highly incompatible.

It was concluded that the fitting method, originally developed and employed in the context of the LOI [196] implementation of the Point 5 analysis does not offer

the stability required for the edge detection. A new and more robust approach is described in the following section.

11.4 The FIR Filter Method for Edge Detection

The new method is inspired by the standard signal processing technique of applying a so-called finite impulse response (FIR) filter to the input data. A detailed description of the method is given in [230], for instance.

11.4.1 Basic Aspects of FIR Filters

An FIR filter of order N_f is a *discrete* and *finite* set of N_f numbers. They can be obtained, for instance, by computing N_f values of a chosen continuous function (e.g., the sine function) on a specific interval (e.g., [-π, π]). The function or expression that is sampled in order to determine the filter values is known as the filter's kernel.

The input data to be analysed is discrete and can also be regarded as a set of N_{data} numbers. The process of applying the FIR filter to the input data is called convolution and it will be described later in this section. A *single* value of the filter response (i.e., the result of one step of the convolution), $R(n)$, is typically calculated as:

$$\begin{aligned} R(n) =& f_0 d(n) + f_1 d(n-1) + ... + f_{N_f} d(n-N_f) \\ =& \sum_{i=0}^{N_f} f_i d(n-i) \end{aligned} \tag{11.25}$$

where f_i represents value i of the filter kernel and $d(n)$ stands for the n-th value of the input data. Typically, the number of filter values N_f is chosen to be considerably smaller than the number of data points.

As a signal processing tool, an FIR filter can be used to reduce the noise that accompanies a measured signal, thus, enabling the studying of the shape, topology and other relevant features of the input data. It can also be used as a means to enhance and identify certain relevant features in the signal data and in this capacity the FIR filter technique is employed in this analysis.

The purpose of this physics study is to identify and localise the edge positions of two energy spectra. In this context, a highly useful and straightforward procedure for edge detection, that consists of applying FIR filters, is offered by the field of computer image processing, as discussed in [231].

In this framework, any given picture containing object edges can be used as input data, for example the image shown in figure 11.9, pannel (a). The basic technique interprets the signal (the input picture) as a two-dimensional function, $f_{ph}(x, y)$, for which the value at any point (x,y) represents the color intensity in that particular pixel. A visible edge is characterised in this case by a high gradient between the values of two neighbouring pixels.

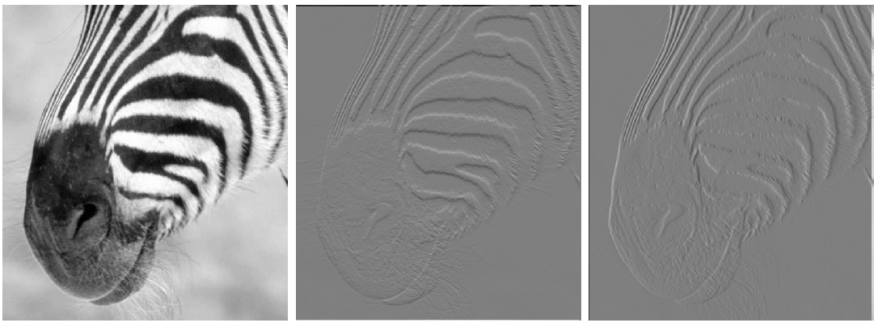


Figure 11.9: Illustration of the edge detection procedure in image processing. Pannel (a) shows the input data, pannel (b) illustrates the result of applying an edge detection filter along the vertical axis and pannel (c) shows the result when the filter is applied along the horizontal direction. Figures taken from [232].

The most straightforward method for detecting the edges in the picture shown in figure 11.9(a) would be to calculate the (first) derivative of $f_{ph}(x, y)$ and observe the location where the derivative provides an extremum.

The calculation of the derivative of the input picture can be approximated by the finite differences method:

$$\begin{aligned} \frac{\partial f}{\partial x} &= \lim_{\epsilon \to 0} \frac{f(x+\epsilon, y) - f(x, y)}{\epsilon} \\ \frac{\partial f}{\partial y} &= \lim_{\epsilon \to 0} \frac{f(x, y+\epsilon) - f(x, y)}{\epsilon} \end{aligned} \tag{11.26}$$

However, while the finite differences method is a very practical way to approximate the first derivative, the outcome is rather dominated by noise.

An alternative approach involves the convolution of the input data, $f_{ph}(x, y)$, with an FIR filter. The FIR filters behave, essentially, in an analogous way to the first derivative, however, the main difference is that they are less sensitive to the background noise.

The application of an FIR filter is demonstrated in the last two pannels of figure 11.9. In this example, the kernel was defined as the first derivative of a Gaussian, where the σ of the Gaussian was set to σ=1 [232]. Pannel (b) shows the convolution with the filter in the y direction, hence, the horizontal stripes are more prominent than the vertical ones. The other case, i.e., the effect of the convolution in the x direction, can be seen in pannel (c), where the vertical stripes are more clearly visible.

In the context of image processing, many different FIR filter kernels, that are analogous to the first derivative approach, have been defined and tested: for example, the Roberts, Sobel [233] or Prewitt operators [234] have been optimised in order to improve the filter response.

11.4.2 The Canny Filter and Its Approximation

In the fundamental paper by J. Canny, [235], a set of performance criteria (also known as the Canny criteria) were mathematically formulated, calculated and applied to several filters in order to determine the optimal kernel for edge detection. The study presented in [235] focussed on one-dimensional edges (as the ones present in the di-jet energy spectra) and, implicitly, one dimensional filters.

According to Canny's theory, the following requirements must be met by a well defined kernel: *(i)* it must maximise the signal-to-noise ratio, *(ii)* the distance Δx between the real and the detected edge must tend to zero, *(iii)* low response multiplicity which means that one real edge should not give rise to more than one response.

In [235], a filter kernel that fulfills all the performance requirements was designed (known as the "Canny filter"). Thus, for the interval [-V, +V], the following expression can be applied:

$$f(x) = a_1 e^{\alpha x} sin(\omega x) + a_2 e^{\alpha x} cos(\omega x) + a_3 e^{-\alpha x} sin(\omega x) + a_4 e^{-\alpha x} cos(\omega x) + c \quad (11.27)$$

where the constants a_1, a_2, a_3, a_4 can be determined from the boundary conditions: $f(0) = f(-V) = f'(-V) = 0$ and α and ω are parameters of the scale and relative weight of the three performance criteria mentioned above.

Clearly, the computation of such a kernel is tedious and highly involved. The first derivative of a Gaussian (FDOG) was proposed in [235] as a simpler and more time-efficient solution. The FDOG function is shown in figure 11.10:

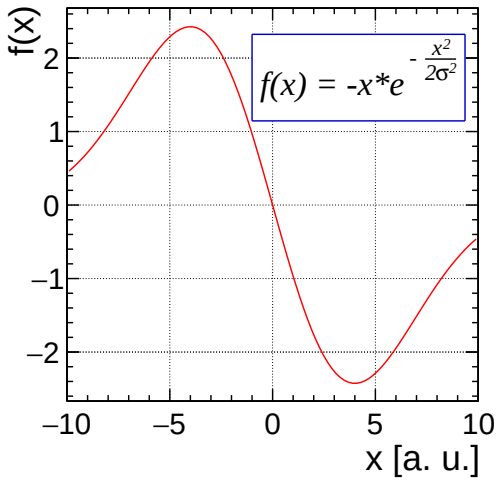


Figure 11.10: The first derivative of a Gaussian function is illustrated using a set of arbitrarily chosen values. It was proposed in [235] as a practical approximation of the Canny filter (equation 11.27).

Since the FDOG function is an approximation of the optimal Canny kernel, it was shown in [235] that, in comparison, its response multiplicity is worse by ≈10%. Nevertheless, despite this performance deterioration, the FDOG function is a close approximation of the Canny filter and is one of the most widely used

kernels due to its straight-forward implementation. Consequently, it was chosen as kernel for the FIR filter implementation used in this thesis.

Example: Applying the FDOG-based Filter on a Step Edge

In order to illustrate the application of the FDOG-based filter the more straight-forward case of a step edge was considered. The edge, shown in figure 11.11, pannel (a), is essentially the output of a Heaviside function for which the discontinuity was arbitrarily set at $x = 10$.

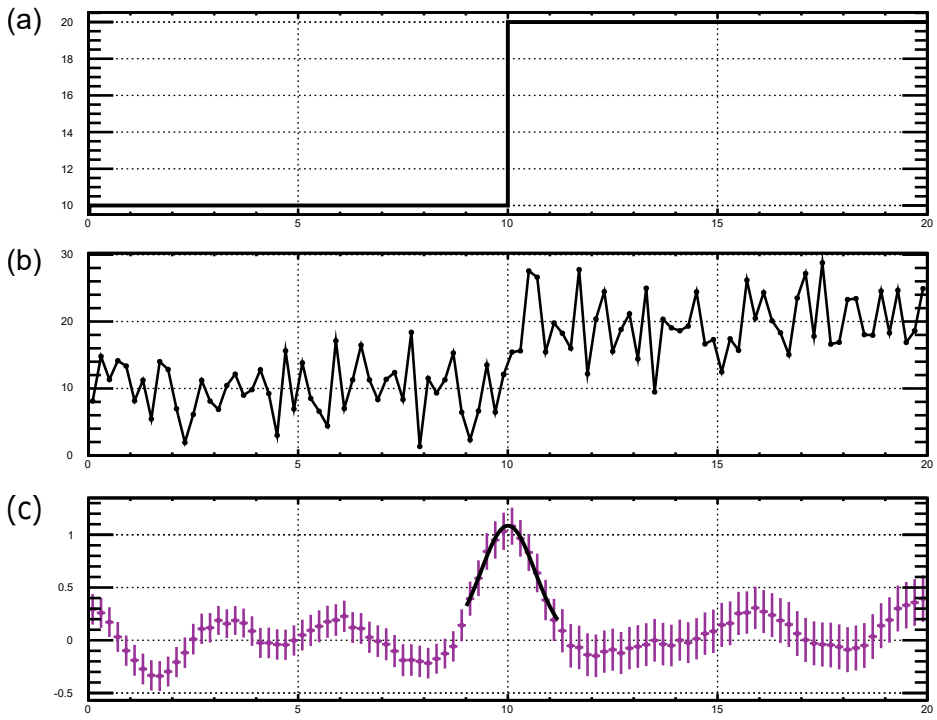


Figure 11.11: Illustration of applying an FIR filter with and FDOG-based kernel on a noisy step edge.

→ **Input Data**: Firstly, a Gaussian noise was added on top of the step edge (pannel (a)), in order to emulate the presence of a background. The obtained noisy edge can be seen in pannel (b) of figure 11.11.
In order to obtain the input (discrete) data on which the FIR filter should be applied, in this example, the function describing the step edge with the additional Gaussian noise was sampled[2] in bins of 0.2 arbitrary units along the x-axis, covering its entire range $(0, 20)$, as shown in figure 11.11(b). In the sampling, the centre of the bin is taken as input x value for the function. Consequently, the obtained set of input data, $d(n)$, contains N_{data}=100 elements: $d(n) = \{d(1), d(2), \ldots, d(100)\}$.

[2] In the case of the di-jet energy spectra, this sampling step is clearly not required: the input data is a histogram, i.e., discrete by definition. This was only necessary for the simplified example.

→ **Kernel coefficients**: The chosen FDOG kernel is defined by the equation $f_{FDOG}(x) = -x \cdot e^{-\frac{x^2}{2\sigma^2}}$, where σ is a tunable parameter. For the present example, a value of σ=4 was arbitrarily chosen. The function was defined over the range of x values $x \in$(-10,10), as shown in figure 11.10.
In order to obtain the (discrete) values of the kernel coefficients, the function $f_{FDOG}(x)$ was also sampled in bins of 0.2 arbitrary units along the x-axis, for convenience. However, the size of the bins used to sample the kernel is a tunable parameter in the FIR filter implementation.
Lastly, as mentioned above, the number of kernel coefficients, N_f, must be smaller than the size of the input data: $\mathrm{N}_f \ll \mathrm{N}_{data}$. In this example, a number of N_f=17 coefficients was chosen.
Thus, the coefficients' values are determined by sampling the $f_{FDOG}(x)$ kernel 17 times, in bins with a width of 0.2 arbitrary units. The sampling is performed such that the central bin corresponds to $x = 0$ and, hence, the coefficient value is $v_0 = f_{FDOG}(0) = 0$. This implies that the remaining 16 coefficient values are obtained by calculating $f_{FDOG}(x)$ for 16 x values that are spread symmetrically in both the negative and the positive direction of the x-axis. The set of the kernel coefficients is represented by: $v(\mathrm{N}_f) = \{-v_k, -v_{(k-1)}, \ldots, -v_1, v_0 \equiv 0, v_1, \ldots, v_{k-1}, v_k\}$, where k=$(\mathrm{N}_f - 1)/2$=8 in this example.

→ **Applying the FIR Filter**: The set of kernel coefficients, $v(\mathrm{N}_f)$, is then convoluted with the set of input data, $d(n)$. There is no a priori prescription on how to "position" the (much smaller) set of coefficients with respect to the set of input data. The convention used in this thesis [236] is that the convolution starts by multiplying the central value of $v(\mathrm{N}_f)$, v_0, with the first entry of $d(n)$, i.e., $d(1)$.
The set of kernel coefficients is then "shifted" on a bin-by-bin basis with respect to $d(n)$. This procedure is repeated such that the filter convolution covers the entire content of the input data set. The last step is reached when the central value of $v(\mathrm{N}_f)$, v_0, is multiplied with the last entry in the data set, i.e., $d(\mathrm{N}_{data})$. For each shift a new value resulting from the filter convolution with the input data is obtained. For the present example it is calculated as:

$$R(i) = -v_8 \cdot d(i-8) + \cdots + v_0 \cdot d(i) + \cdots + v_8 \cdot d(i+8) \tag{11.28}$$

→ **The Filter Response**: The outcome of applying the FDOG-based filter on the step edge data is a set of s values: $R_f(s) = \{R(0), R(1), \ldots, R(s)\}$, known as the *filter response*. They are plotted in the last pannel of figure 11.11.
Typically, the presence of an extremum (i.e., a maximum or a minimum, depending on the specifics of the convolution) in the filter response indicates the location of a detected edge.
In this example, the edge position was determined by fitting a Gaussian function to the peak observed in pannel (c) from figure 11.11. The result of the Gaussian fit revealed that the detected edge position is $\mu = 10.0002 \pm 0.0032$. The χ^2/ndf ratio of the Gaussian fit was $\chi^2/ndf = 1.07$.

The edge position determined from the filter response is in very good agreement with the true position of the edge. Consequently, the filter performance is very promising

11.4.3 Optimisation of the Filter Parameters

Three parameters related to the implementation of the FDOG-based filter can be optimised before applying the method on the di-jet energy spectra obtained from the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ candidate events. The three parameters are: *(i)* the σ of the Gaussian function, *(ii)* the number of kernel coefficients, N_f, i.e., how many times the FDOG function is sampled and *(iii)* the bin size used for the input data histogram.

The optimisation study presented in this section relies on a qualitative comparison of the filter response distributions obtained by varying each parameter's values individually. The optimal configuration of parameter values must provide clear and narrow peaks for the extrema that correspond to the edge positions.

For the optimisation study, two new $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ di-jet energy spectra were randomly generated from the original ones with the same procedure that was used in testing the consistency of the edge fitting method (section "Fit Consistency Test" 190). Since they are the noisiest and, hence, most difficult to measure, the upper edges of the two di-jet energy spectra were chosen as study case.

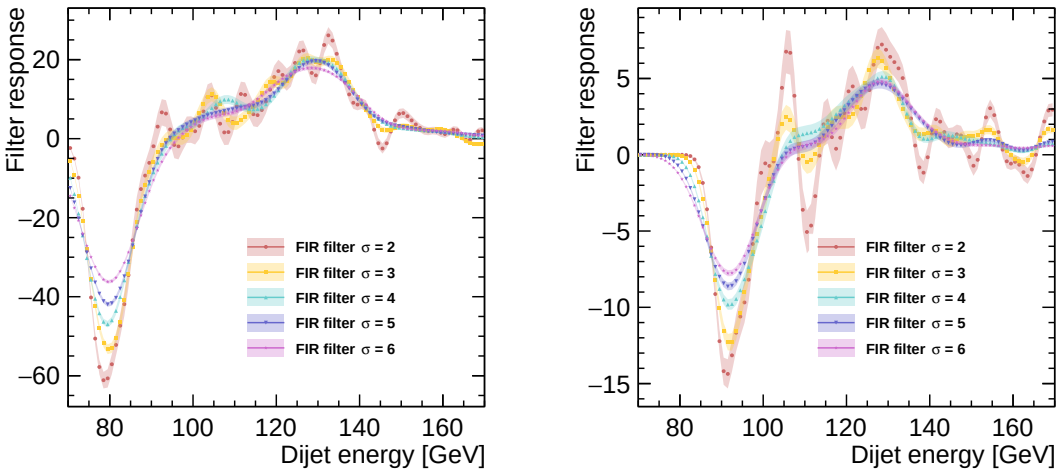
Optimisation of the Gaussian σ:
In order to determine the optimal value of σ, the other two parameters of the filter implementation were fixed. Both input histograms had the same number of bins, $N_{data} = 100$, with the size of 1 GeV per bin. The kernel size was fixed to $N_f = 17$.

The σ values were varied within a range of $\sigma \in \{1 \dots 10\}$ arbitrary units[3] (a.u.), in steps of 1 a.u. The FDOG-based filter was then applied for each σ value on the newly generated $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ di-jet energy spectra.

The very low as well as the very high σ values could be immediately excluded since the corresponding FIR filter response was too noisy for a meaningful evaluation. Figure 11.12 shows the most relevant filter responses and the corresponding σ values.

For both $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ cases, a value of $\sigma = 2$ a.u. was considered too noisy. In contrast, all values above 4 provide almost identical filter responses. Consequently, a value of $\sigma = 4$ a.u. (shown in green in figure 11.12) was chosen as most appropriate for the analysis.

[3] The σ value does not have any specific physical meaning, it is used only for the mathematical computation of the filter kernel, i.e., the first derivative of a Gaussian shown in figure 11.10.



(a) $\tilde{\chi}_1^{\pm}$ candidates **(b)** $\tilde{\chi}_2^0$ candidates

Figure 11.12: Filter response distributions for the optimisation of the Gaussian σ parameter. The kernel size was fixed to $N_f = 17$ and the input histogram bin size corresponds to 1 GeV/bin, i.e., $N_{data} = 100$ (bins).

Optimisation of the kernel size N_f:

In view of the next optimisation step, the σ parameter was fixed to its optimal value, $\sigma = 4$ a.u., while the kernel size was varied. The binning of the input histogram was kept at 1 GeV per bin, i.e., $N_{data} = 100$.

The size of the filter kernel, N_f, is associated with an odd number in this implementation: it consists of an even number of $f_{FDOG}(x)$ values calculated symetrically on both sides of the central value which corresponds to $f_{FDOG}(0) = 0$ and which accounts for the odd final number.

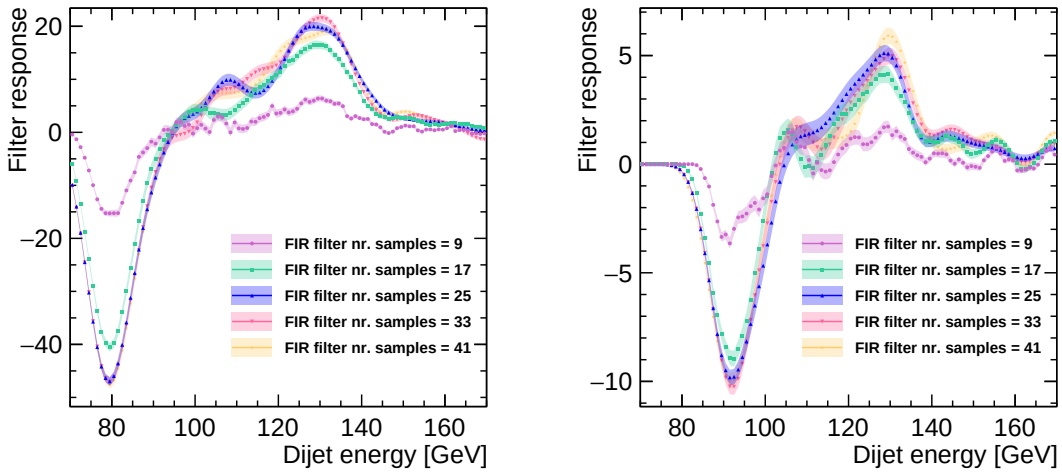
In the present optimisation study, five different kernel sizes from $N_f = 9$ to $N_f = 41$ have been tested. The obtained filter response distributions are shown in figure 11.13.

It was observed that for small kernel sizes like, e.g., $N_f = 9$, the filter will pick up more noise as can be seen in figure 11.13. The width of the peak slightly increases with the size of the kernel, however the peak position does not fluctuate significantly. The narrowest peak was obtained with the kernel size of $N_f = 17$ coefficients (shown in green in figure 11.13), hence, this size was chosen for the final edge detection.

Optimisation of the bin size:

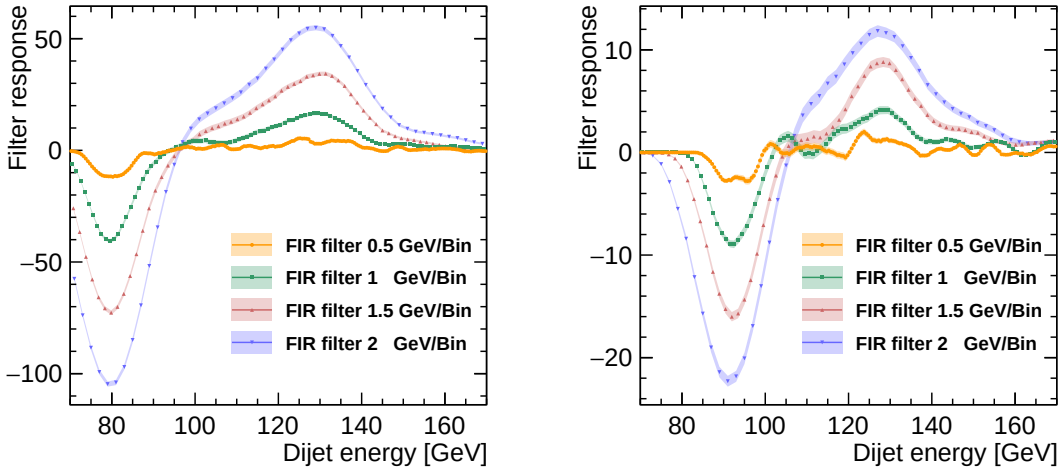
Lastly, in order to find the optimal bin size for the input di-jet energy spectra, the σ parameter and the kernel size were fixed to their previously determined optimal values: $\sigma = 4$ a.u. and $N_f = 17$, respectively. A total of four different bin sizes were considered and the filter response obtained for each of them can be seen in figure 11.14.

It was found that with a binning of 0.5 GeV/bin of the input data, the filter



(a) $\tilde{\chi}_1^{\pm}$ candidates **(b)** $\tilde{\chi}_2^0$ candidates

Figure 11.13: Filter response distributions obtained by varying the kernel size, N_f, while the σ parameter was fixed to $\sigma = 4$ a.u. and the binning of the di-jet energy spectra was set to 1 GeV/bin.



(a) $\tilde{\chi}_1^{\pm}$ candidates **(b)** $\tilde{\chi}_2^0$ candidates

Figure 11.14: Filter response distributions for various bin sizes of the input data histograms. For this evaluation, the σ parameter was fixed to $\sigma = 4$ a.u. while the kernel size was set to $N_f = 17$.

response could not provide clearly distinguishable and relevant maxima due to too much noise (figure 11.14). For bin sizes larger than 1 GeV/Bin the peaks marking the edge positions became wider, however, without significantly altering the edge position itself. Consequently, it was concluded that a data binning of 1 GeV/bin was the optimal choice.

11.4.4 Results: Measured Edge Values

After the parameter optimisation process described above, the FDOG-based filter was applied on the di-jet energy spectra obtained from the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ candidate samples. The energy spectra can be seen in the upper half of the four pannels shown in figure 11.15.

For all four applications, the filter parameters were fixed to their determined optimal values: $\sigma = 4\,\text{a.u.}$ and $N_f = 17$. The input data histograms had a bin size of $1\,\text{GeV/bin}$.

The obtained filter responses[4] are presented in the lower half of the plots. The edge positions are indicated by the peaks in the filter response. Their values were obtained by fitting each maximum with a Gaussian function.

The extracted edge positions, i.e., the mean values obtained from the Gaussian fits, are presented in the fit results labels and are also illustrated by the green dashed lines shown in figure 11.15.

For comparison, the edge positions obtained with the fitting method applied on the DBD data are also presented in the form of the magenta dashed-dotted lines.

Lastly, for completion, the results obtained with the fitting method on LOI data in the original Point 5 study, presented in the ILD *Letter of Intent* document [109, 196], are also indicated by the blue, continuous lines.

It can be seen that the measured lower edge positions are similar for both the $\tilde{\chi}_1^\pm$ and the $\tilde{\chi}_2^0$ case.

Most importantly, it can be observed, from figure 11.15d, that the large discrepancy between the LOI and the DBD fitted $\tilde{\chi}_2^0$ upper edge (table 11.2) was remedied. The initial difference had been caused by the fit (equation 11.23) wrongly identifying the inflection point where the SM contribution becomes dominant as the DBD edge position. This error could be averted by the application of the FIR filter which is more robust with respect to the inherent small fluctuations in the di-jet energy spectra. Thus, it could be concluded that the $\tilde{\chi}_2^0$ high edge identification improved significantly with the use of the FIR filter.

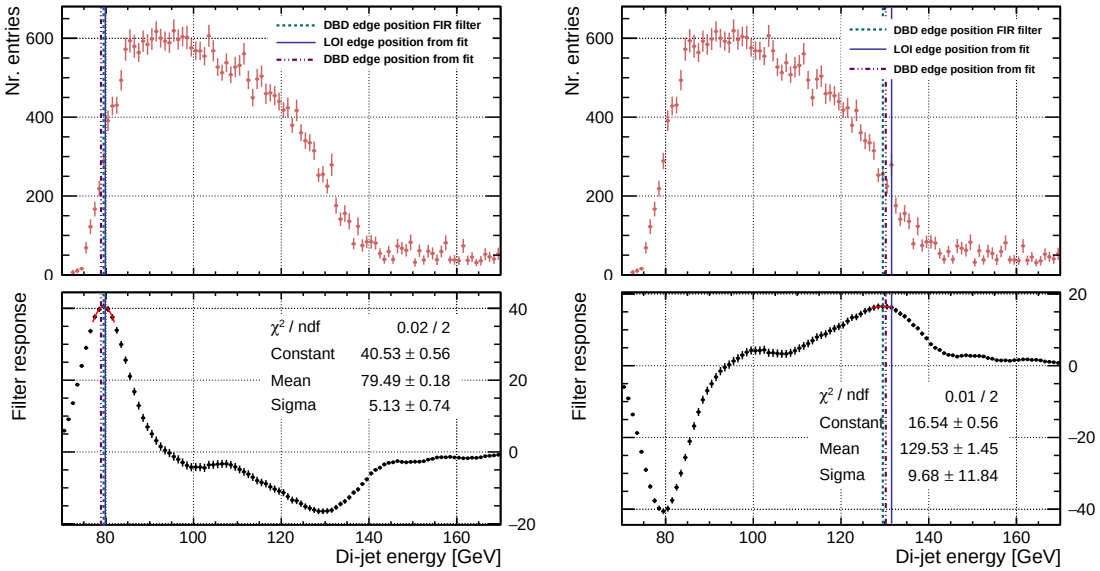
Statistical Uncertainty

The statistical uncertainty of the FIR filter method for edge detection was evaluated in a toy Monte Carlo study.

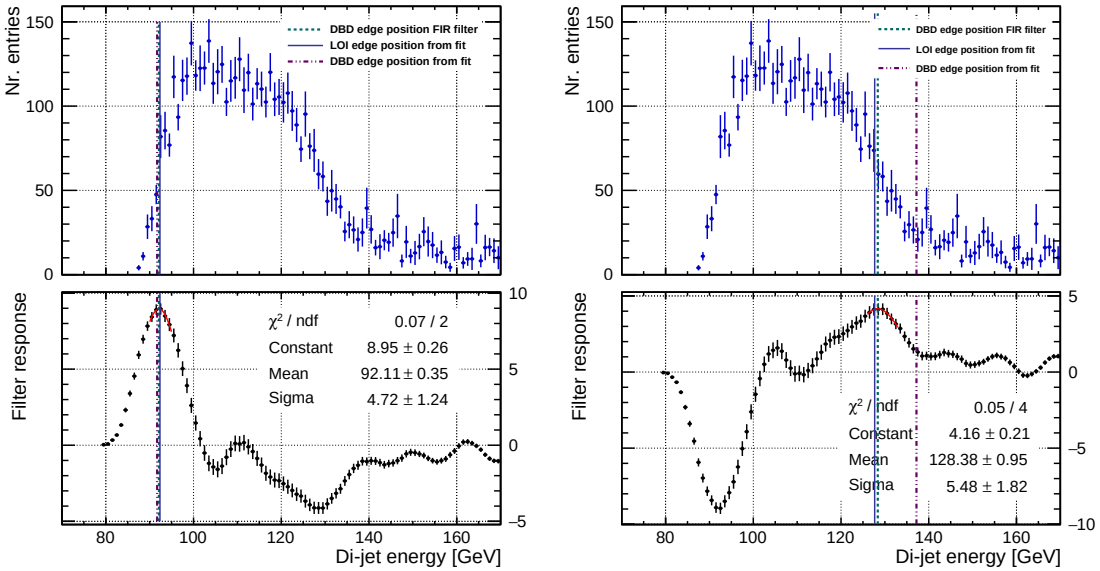
For this purpose, 10^4 new, independent di-jet energy spectra were randomly produced for both the $\tilde{\chi}_1^\pm$ and the $\tilde{\chi}_2^0$ candidate samples. The same randomising procedure based on the original distributions was used, as described in section "Fit Consistency Test" (p. 190).

The FIR filter utilising an FDOG kernel with optimised parameter values, i.e., $\sigma = 4\,\text{a.u.}$ and a kernel size of $N_f = 17$, was applied on each of the 10^4 newly obtained di-jet energy distributions. For each filter output, both the low and the

[4] The filter was inverted for the determination of the upper edges (lower pannels in figures 11.15b and 11.15d). This was done only for convenience, i.e., such that the *descending* lower edge can also be visualised and fitted as a maximum instead of a minimum (as it is shown in figures 11.15a and 11.15c). The filter responses are otherwise identical.



(a) $\tilde{\chi}_1^{\pm}$ candidates (low edge) **(b)** $\tilde{\chi}_1^{\pm}$ candidates (high edge)



(c) $\tilde{\chi}_2^0$ candidates (low edge) **(d)** $\tilde{\chi}_2^0$ candidates (high edge)

Figure 11.15: Results of applying an FDOG-based FIR filter with $\sigma = 4$ a.u. and $N_f = 17$ on the di-jet energy spectra obtained from the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ candidate events. The input histograms are characterised by a bin size of 1 GeV/bin. Both the upper and the lower edges are indicated by the peaks in the filter response.

high edge positions were determined by performing a Gaussian fit to the maxima present in the filter response. Consequently, a distribution of measured edge values is obtained for each of the four edges.

These distributions are shown in figure 11.16. Their widths estimate the statistical precision of the FIR edge extraction method. Therefore, each of the four distributions from figure 11.16 was fitted with a Gaussian function from which

the relevant σ was extracted.

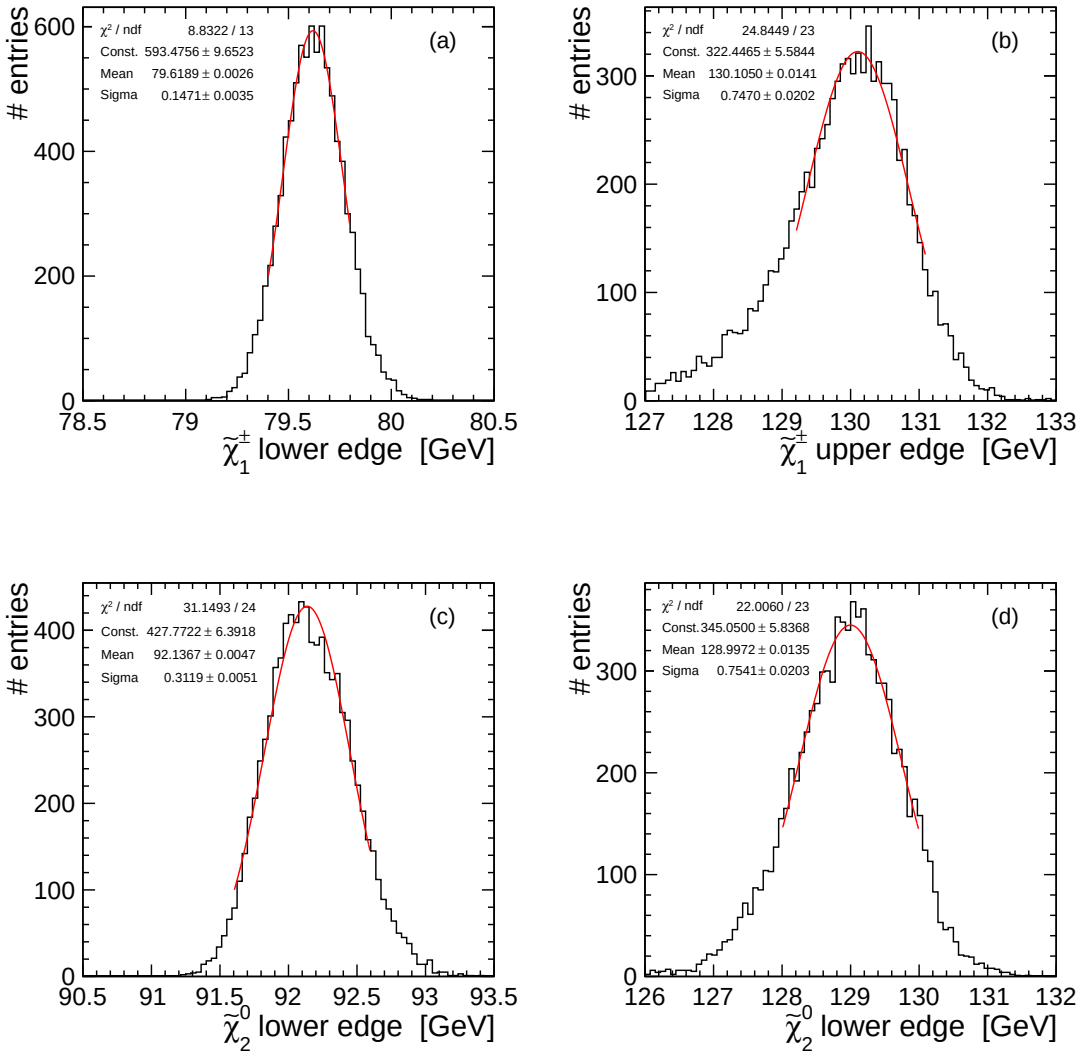


Figure 11.16: The result of the toy Monte Carlo study for determining the statistical uncertainty of the FIR filter method for edge detection.

The obtained results are summarised in table 11.3. The initial results, achieved in the initial Point 5 study performed on the LOI data set and presented in the ILD *Letter of Intent* [109, 196], are also shown.

It can be observed that the FIR filter method for edge identification provides a similar or even lower statistical uncertainty than the fitting method. Moreover, it ensures a much better consistency and robustness against intrinsic small fluctuations in the di-jet energy spectra.

	$\tilde{\chi}_1^{\pm}$ low	$\tilde{\chi}_1^{\pm}$ high	$\tilde{\chi}_2^0$ low	$\tilde{\chi}_2^0$ high
DBD FIR filter	79.49 ± 0.15	129.53 ± 0.74	92.11 ± 0.31	128.38 ± 0.75
$\text{LOI}_{\text{orig.}}$ edge fit	79.88 ± 0.19	131.49 ± 0.74	92.34 ± 0.44	127.67 ± 0.76

Table 11.3: Results of the toy Monte Carlo study performed to evaluate the statistical uncertainty of the FIR filter method for edge identification. The shown edge values were obtained using the DBD data set. The results presented in the initial Point 5 study [109, 196], achieved with the edge fitting method applied on the LOI data set, are shown again for comparison.

Error Propagation

The mass calculation presented in section 11.1 does not account for the unavoidable effects of the beam energy spectrum and gauge bosons' natural widths on the position of the edges. This implies that no analytical expressions can be easily formulated to perform the error propagation and estimate the uncertainty on the determined gaugino masses stemming from these effects.

Consequently, the error propagation was performed based on the toy Monte Carlo study presented above. For each set of measured edge positions (figure 11.16), a new set of gaugino masses was calculated using the equations presented in section 11.1. Thus, for each gaugino, a distribution of 10^4 calculated mass values was obtained.

The average gaugino mass and the corresponding statistical uncertainty were determined by performing a Gaussian fit to the respective calculated mass distribution. The results are shown in table 11.4. The Point 5 model values of the masses are also shown for comparison.

Mass [GeV]	$\tilde{\chi}_1^{\pm}$	$\tilde{\chi}_2^0$	$\tilde{\chi}_1^0$
Model values	216.5	216.7	115.7
Measured values	216.7±3.1	220.4±1.3	118.1±0.9

Table 11.4: Results of the toy Monte Carlo estimation of the uncertainty on the gaugino mass determination.

It can be seen that for $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$, which are both calculated using only the edge positions of the Z boson di-jet energy spectrum, the obtained mass values are 3-4 GeV higher than the model values. The $\tilde{\chi}_1^{\pm}$ mass, which is calculated using the LSP mass and the upper edge position of the W boson energy spectrum, is closer to the model values, but has a significantly larger uncertainty.

11.5 The Edge and Mass Calibration

As mentioned above, the gaugino mass calculation presented in section 11.1 entails some approximations: the effects on the di-jet energy spectra edge positions arising from the gauge bosons' natural widths, the initial state radiation (ISR) and the beam energy spectrum are not taken into account.

In order to account for these approximations and also for the reconstruction effects that cause imprecision in the gaugino mass determination, a mass calibra-

tion procedure can be performed. However, the gaugino masses are not measured directly. In this study, the edge positions are the observables determined directly from the data, hence, they are the ones that must be calibrated.

Firstly, new sets of Monte Carlo data must be produced such that several di-jet energy spectra with different edge position values (on the Monte Carlo truth level) are available. In the next step, the corresponding *measured* edge positions are obtained by applying the FIR filter method on each one of the newly simulated and reconstructed data sets. Finally, the calibration line is determined by plotting the measured edge values with respect to the (known) input ones. The *calibrated* edge values can then be read from the obtained calibration lines.

New Monte Carlo Data Samples

Five new Point 5 SUSY Monte Carlo samples were generated with `Whizard v1.95` [167] and were then simulated and reconstructed with the `DBD` version of the ILD full simulation.

In producing the new data samples, the mass of the lightest supersymmetric particle (LSP), $\tilde{\chi}_1^0$, was left unchanged and set to the model value of $M_{\tilde{\chi}_1^0} = 115.7\,\text{GeV}$. The masses of the two gauginos, $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$, which are almost identical in the Point 5 scenario, were then varied simultaneously.

In order to determine how the input $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ masses should be varied, equations 11.4 and 11.5 were considered. The relation between the gaugino masses and the edge positions can be visualised, as shown in figure 11.17.

Due to the presence of the squared quantities in equation 11.4, the dependency of the edge values on the input masses is a parabola. Consequently, it is possible to have the same edge position value associated with two different input masses as illustrated by the dashed horizontal lines in figure 11.17. However, considering the high and the lower edge *simultaneously* it can be seen that the mass values corresponding to the other ("wrong") edge values do not coincide, e.g., $\mathrm{M_A} \neq \mathrm{M_B}$ in figure 11.17.

In order to avoid this situation, the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ masses were varied both simultaneously from 210 GeV to 225 GeV (indicated by the yellow vertical lines in figure 11.17) in steps of 3 GeV. This ensures that all the new edges fall on the same parabola branch as the original Point 5 values (the red vertical line in figure 11.17).

The original Point 5 model sample will be referred to from now on by "M216" while all the five newly produced mass points will be mentioned with the abbreviation "Mm" with $m \in \{210, 213, 219, 222, 225\}$.

Since the LSP mass is not known a priori, a further calibration procedure should involve fixing the two $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ masses to their Point 5 model values while the $\tilde{\chi}_1^0$ mass is varied instead. However, this was not performed in the present analysis and remains the subject of future studies.

The Standard Model background events produced with the `DBD` version of the full simulation were left unchanged.

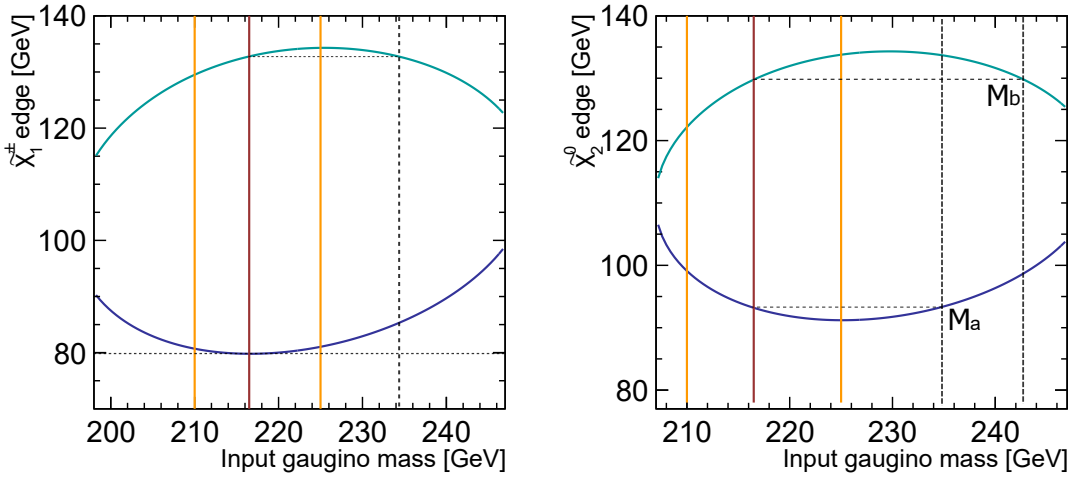


Figure 11.17: The edge positions calculated from equations 11.4 and 11.6 as a function of the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ masses, while the LSP mass was fixed to its nominal value, $M_{\tilde{\chi}_1^0} = 115.7\,\text{GeV}$. The green curves illustrate the high edges while the blue ones correspond to the lower edges.

Extracting the edge positions

In order to avoid making use of any cross-section related information (which may not be available a priori) the number of events produced in each individual mass sample, Mm, was normalised to the number of events that survive the selection cuts and the sample separation in M216. This also ensures that the FIR filter edge detection procedure is applied in the same conditions.

Each one of the five normalised Mm samples is then added, correspondingly, to the unchanged set of SM background $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ candidate events.

The FDOG-based FIR filter with its optimised parameter values, i.e., $\sigma = 4$ and kernel size $N_s = 17$ bins was then run on each total (SUSY and SM) di-jet energy distribution. The edge positions were then extracted from a Gaussian fit to the maxima in the filter response functions.

The Calibration Lines

The measured edge positions, obtained from the FIR filter application on each one of the Mm samples, can then be plotted with respect to the input edge values, thus, producing the required calibration lines.

The purpose of the calibration study performed in this thesis is twofold. Firstly, the calibrated edge positions are to be determined and, based on their values, the (calibrated) gaugino masses are calculated. Furthermore, depending on the Monte Carlo data production stage at which the edge values are calibrated, different effects that occur at that level can be investigated, as described in the following:

→ Generator level versus calculated edges

In order to investigate the effects of the boson natural width, the beam energy spectrum and ISR emission on the edge positions, their values measured on *gen-*

erator level can be calibrated with respect to the positions calculated with the formulae 11.4 and 11.5 (which do not take these effects into account).

It must be noted that no particle-detector interaction is simulated at the generator stage. Consequently, the standard reconstruction and data preparation tools used in the analysis (chapter 10) were not applied. For instance, no jet clustering was performed. The di-jet energy spectra were obtained directly from the Monte Carlo truth information concerning the energy of the quark pairs.

Furthermore, the observables used in the event selection and in the signal separation are not defined on the generator level. Therefore, it is not feasible to select $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ candidate events in this context. In order to circumvent this issue, the Standard Model and SUSY backgrounds were not considered in the study.

The edge detection was performed by applying the FIR filter directly on the pure (i.e., Monte Carlo true) chargino and neutralino signal di-jet energy distributions. This was carried out for each one of the five newly produced "Mm" samples, with $m \in \{210, 213, 219, 222, 225\}$.

Since the $\tilde{\chi}_1^{\pm}$ low edge overlaps with the mass of the W boson, it is not very useful in the mass calculation. Consequently, it was neglected in the calibration procedure.

Thus, for each mass sample, Mm, three values, corresponding to the $\tilde{\chi}_1^{\pm}$ high edge and the two $\tilde{\chi}_2^0$ edges, were obtained with the FIR filter method.

In order to obtain the calibration lines, each edge was considered individually. The five measured values (each one determined from an Mm sample) were plotted with respect to the corresponding calculated values. The results are presented in figure 11.18.

The calibration lines, shown in red, consist of a linear fit applied to the five points of the graph. In order to avoid any bias, the data point belonging to M216 was completely excluded.

The slope of the chargino upper edge calibration curve is very close to one, i.e., the ideal case. This would imply that the effects considered in the calibration have a small impact on the edge positions.

However, in the case of the neutralino edges, the two slopes have a value of ≈0.8. This is due to the fact that, for M210, the phase space for the reaction $\tilde{\chi}_2^0 \to Z\tilde{\chi}_1^0$ tends to zero since $M_{\tilde{\chi}_2^0} - M_{\tilde{\chi}_1^0} \approx 95\,\text{GeV}$. Furthermore, since the phase space is small, the influence of the beam energy spectrum and of the gauge boson width is larger.

To determine a calibrated edge position, the measured M216 value was projected onto the linear fit (dashed line in figure 11.18). The intersection point with the calibration function was then further projected onto the x axis. This final projection provides the calibrated value which can be read directly off the horizontal axis. The uncertainty on the measured values was also taken into account in the procedure. Thus, in order to calculate the calibrated error, the same projections were applied for the uncertainty intervals as well. Table 11.5 summarises the calibration results.

From the values in table 11.5, the effect of the beam spectrum, ISR and natural boson width can be estimated. Thus, comparing the "Generator" and the

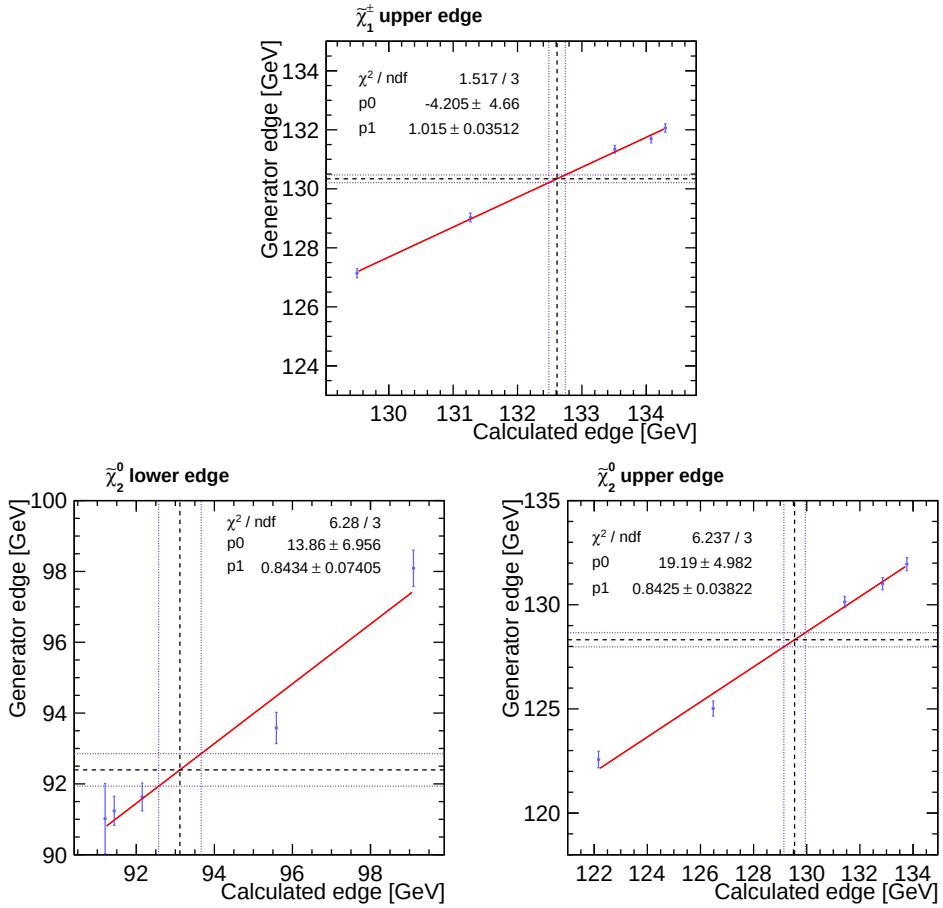


Figure 11.18: The generator level edge positions versus the calculated edge positions.

	$\tilde{\chi}_1^{\pm}$ (high) (GeV)	$\tilde{\chi}_2^0$ (low) (GeV)	$\tilde{\chi}_2^0$ (high) (GeV)
Generator	130.34±0.17	92.39±0.46	128.32±0.34
Calculated	132.76	93.09	129.92
Calibrated	132.61±0.13	93.11±0.55	129.54±0.4

Table 11.5: Results of the edge calibration comparing generator level versus calculated edges.

"Calculated" columns, the measured edges appear to be shifted towards smaller values by: 1.8% in the case of the chargino higher edge, 0.8% for the neutralino lower edge and 1.2% for the upper edge of the neutralino. It can be seen that the edge values measured with the FIR filter are incompatible with the calculated edge values by up to $\approx 7\sigma$, on average. However, as can be seen from the third column, the calibration procedure brings the generator level edge values closer to the calculated ones, as expected.

Furthermore, since the slopes of the $\tilde{\chi}_2^0$ calibration curves are smaller than one, the calibrated uncertainties on the edge positions are slightly higher: by 19.56% for the neutralino lower edge and 17.65% for the higher edge.

Once the calibrated edge values and uncertainties were obtained, the gaugino masses were re-calculated with the formulae discussed in 11.1. A toy Monte Carlo experiment was performed in order to perform the error propagation. The calibrated edge positions were varied within their newly determined calibrated uncertainties and with each randomised set of edge values the gaugino masses were then computed. The procedure was repeated 10^6 times.

The impact of the edge calibration on the gaugino mass calculation on generator level is summarised in table 11.6.

Gaugino	Mass [GeV] (before calib.)	Mass [GeV] (after calib.)	Model mass [GeV]
$\tilde{\chi}_1^{\pm}$	217.92±2.49	216.57±2.74	216.5
$\tilde{\chi}_2^0$	219.65±1.54	216.74±1.54	216.7
$\tilde{\chi}_1^0$	118.195±0.78	115.99±0.85	115.7

Table 11.6: Comparison of mass values before and after the calibration procedure. The Point 5 model masses are also shown for comparison.

It is clearly visible that the edge calibration contributes to bringing the M216 gaugino masses to their initial model values - in other words to counteract the beam spectrum, ISR and gauge boson width effects. Thus, the calibration improves the measured masses by 0.12% for the $\tilde{\chi}_1^{\pm}$ case, by 1.03% for the $\tilde{\chi}_2^0$ and by 1.31% for the LSP mass.

The fact that the calibrated masses are close and compatible within their uncertainties to the model values offers a confirmation that the calibration procedure was succesful and that the FIR filter's performance is robust and consistent, independently of the assumed SUSY mass model.

→ Reconstruction level versus generator level edges

In order to investigate the impact of the reconstruction on the edge positions and implicitly on the measured gaugino masses, another calibration study was performed. For this purpose, the reconstructed edge positions were plotted versus the generator level edges, as shown in figure 11.19.

As mentioned earlier, the selection and sample separation procedures cannot be applied on the data at the generator level. Since the effects of the background contributions were not of interest to this study, it was decided to consider only the chargino and neutralino signal samples.

In order to make a correct and realistic comparison, the background samples were also removed from the *reconstructed* di-jet energy distributions. The $\gamma\gamma \rightarrow$ to hadrons background that is overlaid *after* the generator stage had already been removed when preparing the reconstructed distributions. However, the chargino and neutralino signal samples on the reconstructed level were also affected to a certain percentage by the sample selection and separation processes. Thus, in order to correct for that, those signal events that did not survive the selection cuts and the separation conditions were also removed from the distributions on the generator level. This ensured that the comparison is made between the exact same

events - the only difference between them being their Monte Carlo production stage.

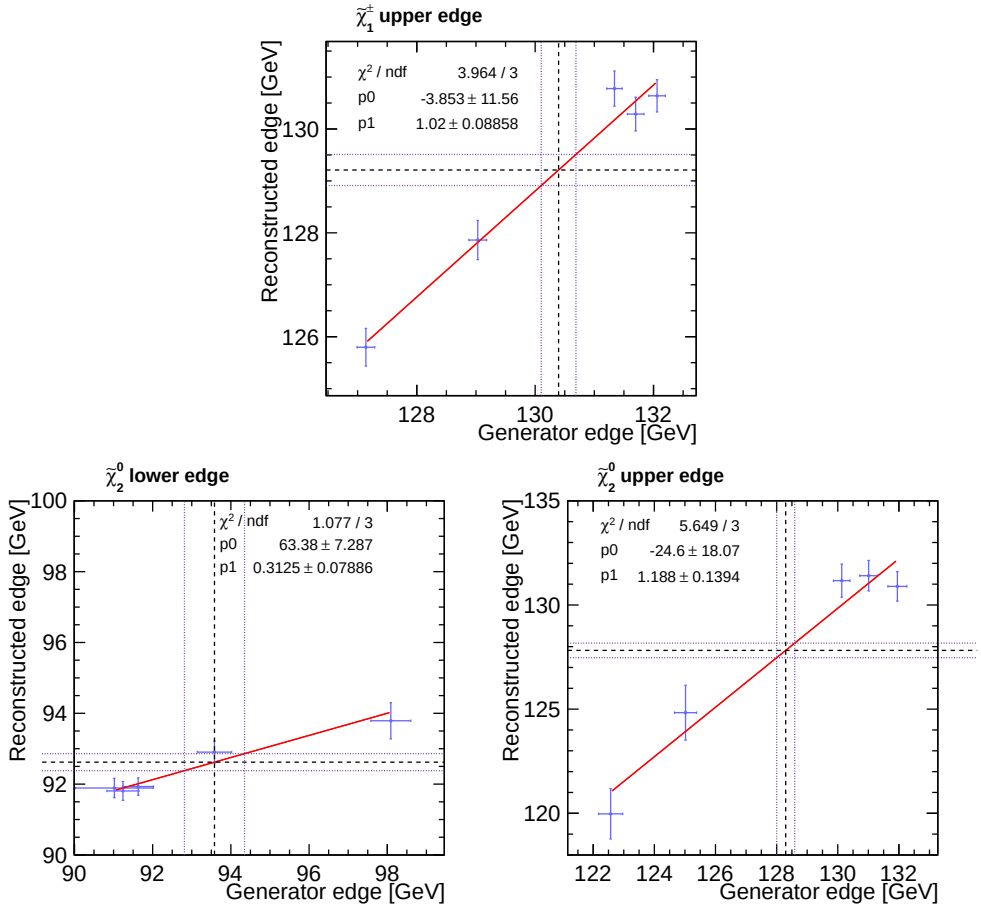


Figure 11.19: The edge position on reconstruction level versus the generator level.

After the sample preparation and the application of the edge detection procedure, the edge calibration was performed. The result can be seen in figure 11.19. The slopes of all three edges are rather close to one which indicates that the reconstruction effects do not have a major impact on the reconstructed measured edges.

→ Reconstruction level versus calculated edges

The impact of the reconstruction effects alone was found to be smaller than the generator level effects in the previous step.

The final and most important ingredient in the mass calibration study was to compare the edges determined from the di-jet distributions on the reconstruction level to the values calculated with the mathematical approximation provided by the formulae 11.4 and 11.5.

Thus, the aim of this final study was to account for the combined impact of the gauge boson width and beam effects together with the reconstruction effects on the measured edge positions. Furthermore, the influence of the SM and SUSY backgrounds contributions was also included.

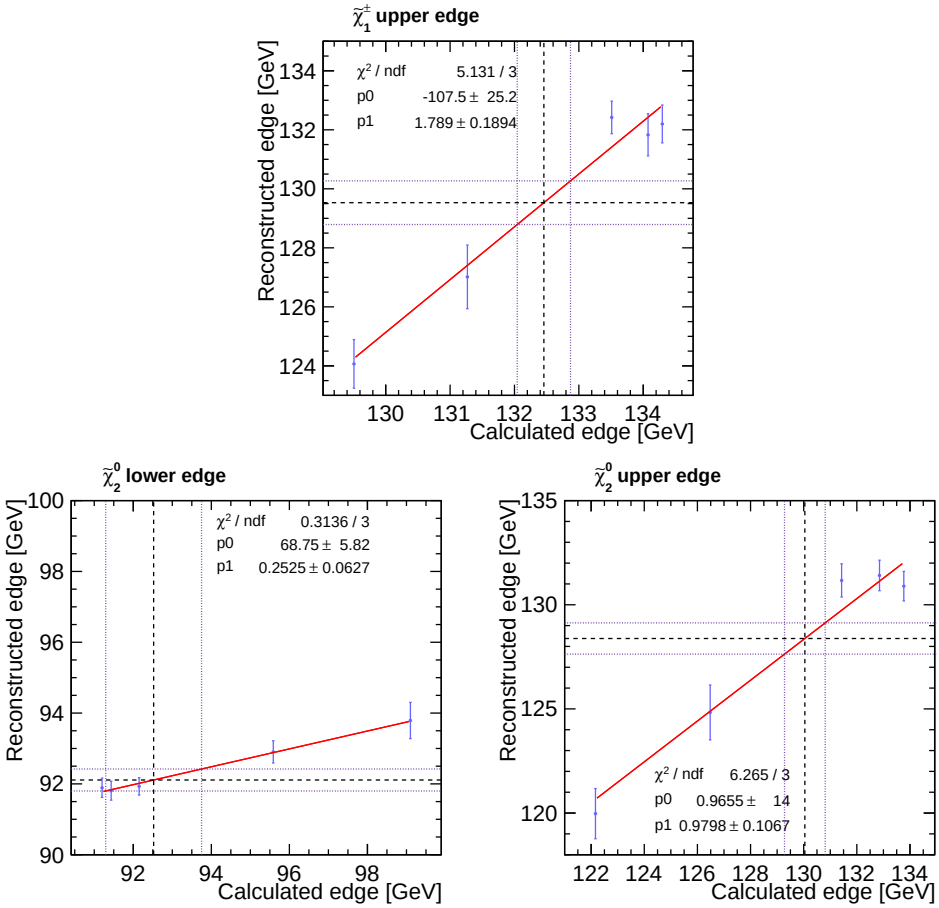


Figure 11.20: The edge position on reconstruction level versus the values calculated with formulae 11.4 and 11.5.

The data samples consisted of all the SUSY and SM events that survived the event selection (chapter 10) and sample separation criteria.

The FIR filter was then applied on the reconstructed di-jet energy samples and the calibration procedure was performed using the measured reconstructed edge positions. The results are illustrated in figure 11.20.

	$\tilde{\chi}_1^{\pm}$ (high) (GeV)	$\tilde{\chi}_2^0$ (low) (GeV)	$\tilde{\chi}_2^0$ (high) (GeV)
Reconstructed	129.53±0.74	92.11±0.31	128.38±0.75
Calculated	132.77	93.09	129.92
Calibrated	132.46±0.44	92.52±1.23	130.04±0.77

Table 11.7: Results of the edge calibration comparing reconstruction level edge values to the calculated values.

The ideal slope of the calibration linear fit has a value equal to one. This would indicate that the effects that were considered in the calibration would have a small impact on the edge positions. In this study, while the neutralino upper edge is

close to the ideal value, figure 11.20 clearly shows a deterioration that can be observed especially in the case of the neutralino lower edge. This is not entirely unexpected since the presence of the SM and SUSY background may be one of the main reasons for the deterioration of the slopes corresponding to the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ calibration curves.

Gaugino	Mass [GeV] (before calib.)	Mass [GeV] (after calib.)	Model mass [GeV]
$\tilde{\chi}_1^{\pm}$	216.7±3.1	214.1±4.8	216.5
$\tilde{\chi}_2^0$	220.4±1.3	216.9±3.4	216.7
$\tilde{\chi}_1^0$	118.1±0.9	115.5±1.8	115.7

Table 11.8: Comparison of mass values before and after the calibration.

Table 11.7 presents the edge calibration results and the comparison between the reconstructed and the calculated edge positions. The cumulative impact of the natural boson width plus the beam and reconstruction effects can be obtained. By comparing the "Reconstructed" and the "Calculated" columns in table 11.7 it can be seen that the chargino upper edge decreased by 2%. In the neutralino case, this effect is of the order of 1.05% for the lower edge while the upper edge decreases by 0.72%.

Lastly, table 11.8 shows the outcome of the calibration procedure when calculating the gaugino masses for M216. While the mean $\tilde{\chi}_1^{\pm}$ is shifted lower by 1.2%, the final measured mass is well compatible with the original model values. The neutralino mass was corrected by 1.6% while the LSP mass was improved by 2.15%. This is, however, achieved at the cost of a significant increase of the uncertainty.

CHAPTER 12

Cross-Section Measurements

Based on the event selection presented in section 10.4, the prospects for measuring the cross-section times branching ratio for the two Point 5 signal processes:

$$\begin{aligned} \sigma(e^-e^+ &\rightarrow \tilde{\chi}_1^-\tilde{\chi}_1^+) \times BR(\tilde{\chi}_1^-\tilde{\chi}_1^+ \rightarrow W^-\tilde{\chi}_1^0 W^+\tilde{\chi}_1^0) \\ \sigma(e^-e^+ &\rightarrow \tilde{\chi}_2^0\tilde{\chi}_2^0) \times BR(\tilde{\chi}_2^0\tilde{\chi}_2^0 \rightarrow Z^0\tilde{\chi}_1^0 Z^+\tilde{\chi}_1^0) \end{aligned} \tag{12.1}$$

will be evaluated and discussed in the following.

As mentioned in chapter 9, in the considered SUSY model, the branching ratios are 99.4% for the $\tilde{\chi}_1^\pm \rightarrow W^\pm\tilde{\chi}_1^0$ decay and 96.4% for the $\tilde{\chi}_2^0 \rightarrow Z^0\tilde{\chi}_1^0$ case. The determination of the branching ratios for these as well as other available decay modes is left for future studies. The term "cross-section" will from now on be employed to denote the measured quantity $\sigma \times BR$ for the remainder of this section.

The basic method for measuring the cross-section relies on the following expression:

$$\sigma = \frac{N_{obs.} - \langle N_{bgrd.} \rangle}{\epsilon \int \mathcal{L} dt} \tag{12.2}$$

where $N_{obs.}$ represents the number of observed events, $\langle N_{bgrd.} \rangle$ is the number of expected background events for the integrated luminosity $\int \mathcal{L} dt$ and ϵ is a measure of the detector, reconstruction and event selection efficiencies.

The cross-section is proportional to the number of observed signal candidate events, as shown in equation 12.2. Since this study considers two signal processes simultaneously, it is therefore essential to determine both the number of WW and the number of ZZ signal-like events present in the final data sample.

In the initial Point 5 study, performed on the LOI data set and published in the ILD *Letter of Intent* [109, 196], two different approaches were proposed for this purpose. In both cases, the relevant observable is the di-jet mass distribution, after applying all the preselection cuts (section 10.4).

The first method comprises a one-dimensional fit to the di-jet mass distribution, *after* the kinematic fit, when only the best di-jet configuration is considered. The second method consists of a two-dimensional template fit performed on the distribution of one di-jet mass versus the other, without the kinematic fit. For the second method, two cases have been studied: *(i)* all possible di-jet configurations are taken into account and *(ii)* similarly to the one-dimensional fit, only the best di-jet configuration is employed. Both strategies rely on the assumption that the Standard Model (SM) di-jet distribution is known or can be obtained from other sources.

In the following sections, the analysis techniques are demonstrated using the DBD Monte Carlo data set. The original (LOI [109, 196]) methodology is followed in general. However, several new approaches aiming at improving the original methods emerged in the context of this thesis. They are described in detail and indicated as such in the following.

12.1 One-Dimensional Boson Mass Fit

As mentioned above, the one-dimensional fit strategy for determining the cross-section makes use of the invariant mass distribution of the di-jet system, after all the selection cuts (Level 2) and the kinematic fit have been applied.

In order to obtain the relevant distribution, as shown in figure 12.1, the selected events were required to pass the kinematic fit convergence test. Furthermore, only one mass value (due to the kinematic fit equal mass constraint) was used to fill the distribution, namely, the mass corresponding to the jet configuration for which the kinematic fit has the best χ^2 value.

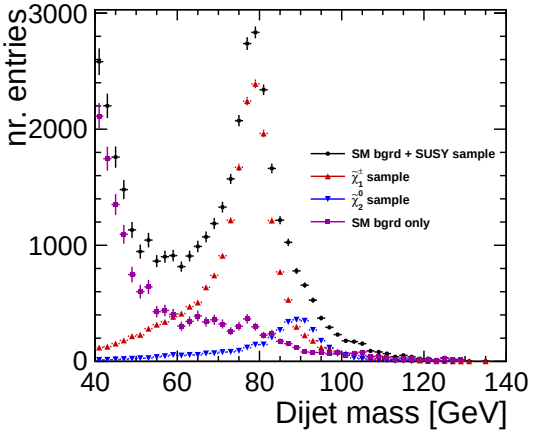


Figure 12.1: The di-jet mass distribution obtained after applying the event selection and the kinematic fit.

The relevant observable, i.e., the di-jet mass, is illustrated in figure 12.1. The total final distribution can be seen in black, while the individual contributions from the Standard Model (SM) background, the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ signal samples can be seen in magenta, red and blue, respectively.

It is relevant to note that, due to the preselection, while the SM background features dominantly in the low di-jet mass region, its contribution in the signal region (≈60 GeV to ≈115 GeV) is small. Furthermore, in contrast to the SM distribution in the signal region, the signal peaks - corresponding to the W and Z boson mass peaks - are Gaussian-like and clearly distinguishable.

The relevant information for the cross-section determination is the number of signal $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ events that can be found in the final sample. In order to extract this information, the total di-jet mass distribution is fitted such that the only free parameters of the fit are essentially the normalisation parameters that describe the dominant chargino and the smaller neutralino peaks. The normalisation parameters provide a direct estimate of the number of signal events that can be found under the signal peaks. The formulation of the fitting function is described in the following.

The Standard Model Background Fit

When constructing the fitting function, the assumption was made that the Standard Model contribution is well known and can be described by a component of the total fitting function.

The same function, i.e., a second degree polynomial, that was employed in the initial LOI analysis [237] was used to fit the SM background, for legacy reasons. The result of the polynomial fit can be be seen in figure 12.2. The ratio of the fit's χ^2 ($\chi^2 = 64.53$) to the number of degrees of freedom ($ndf = 27$) is $\chi^2/ndf = 2.39$.

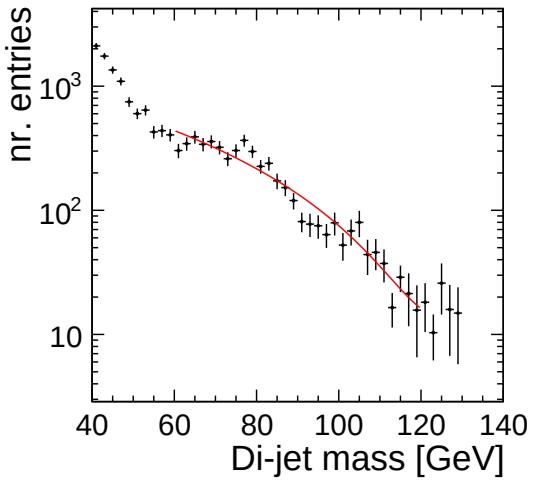


Figure 12.2: The result of fitting the SM contribution to the total di-jet mass distribution with a second order polynomial function.

In contrast to the original LOI result (figure 6 of reference [237]), it can be seen both from the rather large χ^2/ndf value and also from figure 12.2 that the second order polynomial function does not provide the best description of the SM contribution in the DBD data case.

This is not unexpected: for instance, the more realistic DBD version of the ILD

full simulation includes the $\gamma\gamma \rightarrow hadrons$ overlay while the original LoI version did not. The procedure applied to remove it (section 10.2) can affect the di-jet energy and, implicitly, the di-jet mass such that the final SM distribution shown in figure 12.2 differs to some extent from the original LOI one.

Nevertheless, similarly to the original LOI analysis, the values of the polynomial coeffiecients obtained from the fit (figure 12.2) were fixed in the present DBD study as well, for comparison and legacy reasons. The obtained SM fit was then used as such in the total fitting function.

The Total Fitting Function

The two signal $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ contributions can each be described by a non-relativistic Breit-Wigner distribution (traditionally used for fitting resonances) convoluted with a Gaussian distribution which accounts for the detector resolution. This convolution is known as the Voigt function. Consequently, the total fitting function, applied to both signals plus the SM background (black markers in figure 12.1), can be written as:

$$f_{Fit}^{1d}(x) = a \cdot x^2 + b \cdot x + c + N_W \cdot Voigt(x - \mu_W, \sigma_W, \Gamma_W) + N_Z \cdot Voigt(x - \mu_Z, \sigma_Z, \Gamma_Z) \tag{12.3}$$

where N_W and N_Z are the normalisation constants (leading to the number of candidate events) and the only free parameters of the fit. μ_W and μ_Z are the mean mass values of the W and Z bosons. σ_W and σ_Z represent the detector resolution and Γ_W and Γ_Z are the gauge bosons' natural widths.

The goal is to have N_W and N_Z as the only free global fit parameters. Since the a, b and c coefficients are fixed to the values obtained from the SM fit, the parameters of the two Voigt functions are now to be determined and fixed as well.

In the original LOI analysis [237], the μ_W, μ_Z and Γ_W, Γ_Z parameters were all fixed to their nominal values, as quoted by the Particle Data Group (PDG) [13]. This is also the case in the present DBD-based study, but only for the two $\Gamma_W = 2.085\,\text{GeV}$ and $\Gamma_Z = 2.495\,\text{GeV}$ parameters. The rest of the Voigt parameters, i.e., the detector resolution and the boson mass peak values, are to be determined from data.

The advantage of obtaining the boson masses from data instead of using the PDG values is that, this way, other reconstruction related effects - which cannot be accounted for by the Gaussian convolution alone - can thus be taken into account. Considering the final signal states that this study is interested in, the most obvious choice of data sample for extracting the relevant mass and resolution values would be the Standard Model $e^+e^- \rightarrow W^+W^- \rightarrow$ hadrons and $e^+e^- \rightarrow Z^0Z^0 \rightarrow$ hadrons samples. However, several issues arise when considering these data samples.

Firstly, as shown in section 10.1, in the $\tilde{\chi}_1^{\pm}$ signal sample, the W boson mass was inadvertently lowered to $M_W = 79.8\,\text{GeV}$ by the event generator. This feature is present only in the signal sample while the SM $W^+W^- \rightarrow$ hadrons sample was not affected. Therefore, a difference of $\approx$0.6 GeV, on the generator level, must be taken into account when extracting the reconstructed W boson mass from the Standard Model sample and employing it in the global fit.

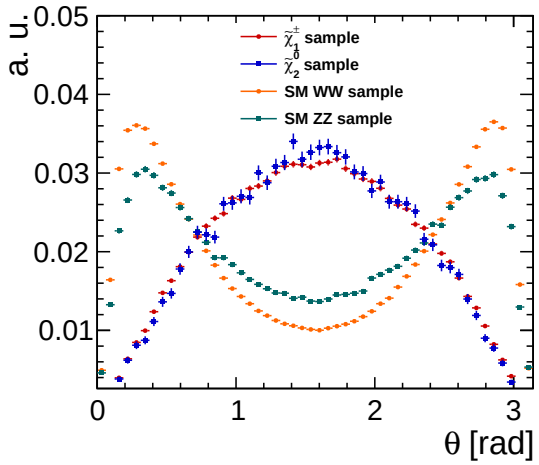


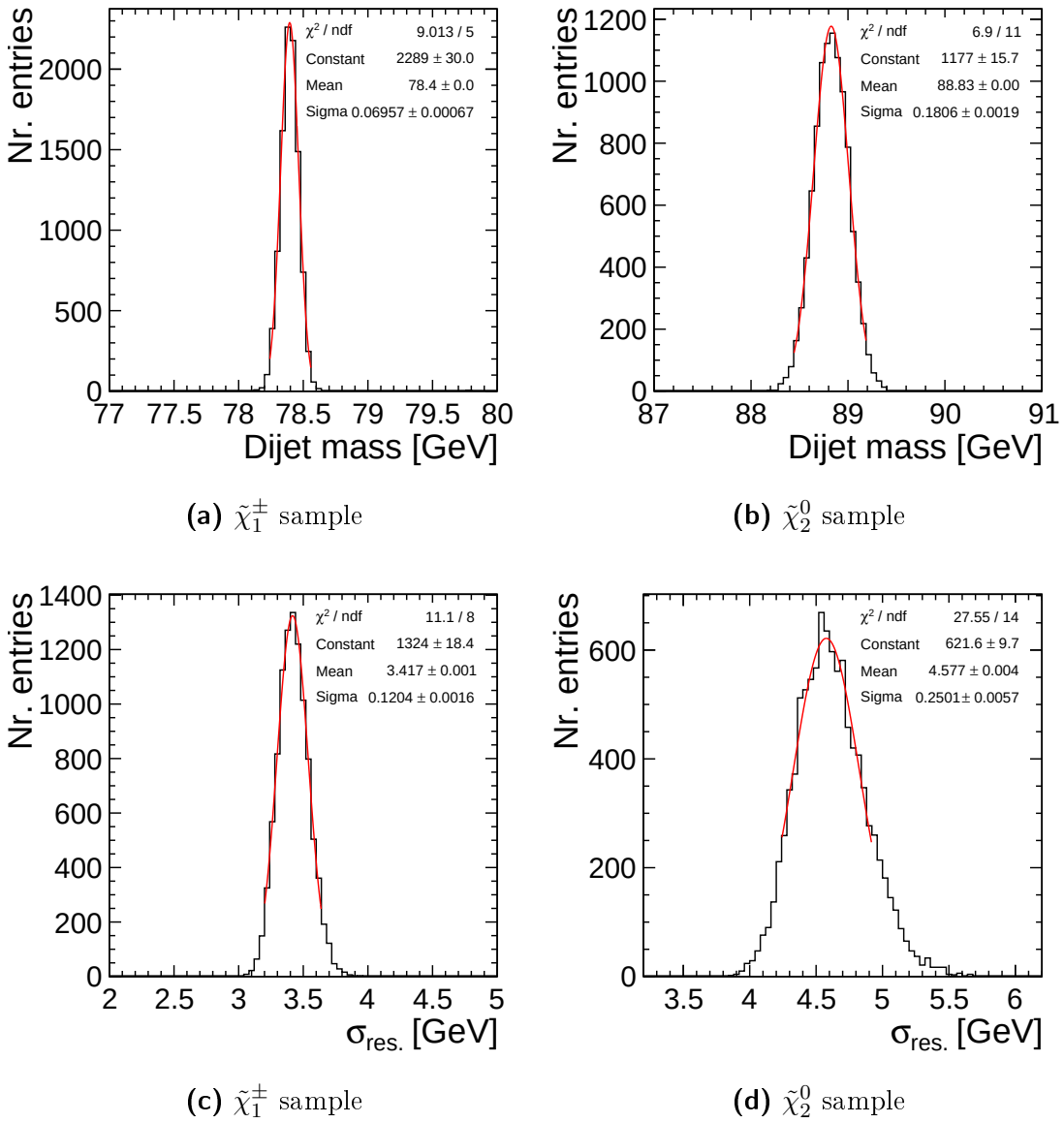
Figure 12.3: Comparison of the jet theta distribution between the Standard Model and SUSY samples of WW → hadrons and ZZ → hadrons.

Furthermore, as shown in figure 12.3, the jet topology of the signal $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ events is very different from that of the Standard Model W^+W^- and Z^0Z^0 samples. While, due to the large amount of missing energy, the signal events are mostly oriented in the central region of the detector, the Standard Model jets are highly boosted and peak in the forward region. Therefore, in order to obtain more meaningful information on the detector resolution, information that could be then input into the global fit, the dependency of the detector resolution on the (jet) θ angle would have to be investigated. However, due to the limited amount of Standard Model Monte Carlo data samples, the resolution $\sigma^{res}(\theta)$ dependency is still to be studied.

In view of all these aspects, a different strategy was considered in this thesis for the determination of the parameters of the two the Voigt functions. This strategy consists of performing a toy Monte Carlo study based on the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ signal samples themselves in order to extract the relevant parameters and will be described in the following paragraphs.

For the toy Monte Carlo procedure, the relevant di-jet mass distributions were produced for each one of the signal samples and each distribution was used as input for the toy Monte Carlo data. Essentially, each di-jet mass histogram was interpreted as a probability density function which was randomly sampled N times in order to produce a new, statistically independent distribution. N represents the number of entries in the new distribution and it was obtained by randomly sampling a Gaussian distribution centred around the number of events of the initial histogram ($N_{inititial}$) and with a width equal to $\sqrt{N_{inititial}}$.

The procedure was repeated 10^4 times, hence, producing 10^4 new independent distributions, for each signal sample. A (single) Voigt function was fitted to each one of the randomly produced distributions and from each fit the relevant parameters ($M_{W/Z}$ and $\sigma^{res}_{W/Z}$) have been extracted. Thus, 10^4 different values have been obtained for each fit parameter of interest. The result of the toy Monte Carlo study is shown in figure 12.4.



(a) $\tilde{\chi}_1^{\pm}$ sample **(b)** $\tilde{\chi}_2^0$ sample

(c) $\tilde{\chi}_1^{\pm}$ sample **(d)** $\tilde{\chi}_2^0$ sample

Figure 12.4: Result of toy Monte Carlo study. The obtained parameter values will be fixed in the global fit.

The average values of the boson masses are shown in figures 12.4a and 12.4b: $\langle M_W \rangle = 78.39 \pm 0.07\,\text{GeV}$ and $\langle M_Z \rangle = 88.83 \pm 0.18\,\text{GeV}$. In both cases, a shift in the average mass values towards lower masses is noticeable. This effect, ranging from 1% to 3%, is observable both on the Standard Model as well as on the SUSY data samples. There are several potential causes for the observed shift, starting from the inevitable initial state radiation (ISR) emission and beam spectrum effects, to the more probable reconstruction issues surrounding the jet clustering algortihm, the removing procedure of the so called $\gamma\gamma$ background (described in section 10.2) which can inadvertently cut out slightly more energy from the event than intended as well as the outcome of the kinematic fitting which does not always lead to the correct jet configuration. One potential solution to correct this

feature would be to raise the jet energy scale by 1% to 3%. However, for the present analysis the mass shift is considered a systematic effect and is treated as such. Thus, the mass values quoted above are fixed with their respective values in the global fit.

The resolution average values are presented in figures 12.4c and 12.4d. The toy Monte Carlo study provides an average value of $\sigma_W^{res} = 3.42 \pm 0.07\,\text{GeV}$ for the chargino sample and $\sigma_Z^{res} = 4.58 \pm 0.25\,\text{GeV}$ for the neutralino case.

The much larger resolution value obtained for the $\tilde{\chi}_2^0$ data is most likely an effect of the much lower statistics available for the $\tilde{\chi}_2^0$ events. This can be seen from figure 12.1 where the $\tilde{\chi}_1^\pm$ peak (i.e., a measure of the number of signal events) is considerably higher with respect to the $\tilde{\chi}_2^0$ one.

Since the toy Monte Carlo study has produced two values for the detector resolution, different by approximately 1 GeV, the global fit has been performed several times with different combinations of the obtained σ^{res} values in order to estimate the impact of the W and Z resolution values on the statistical precision of the one-dimensional fitting method. Thus, the following configurations have been used, in turn, in the global fit.

First, the individual values of σ_W and σ_Z have been used $(\sigma_W^{res}, \sigma_Z^{res})$, as determined directly from the toy Monte Carlo. Figure 12.5 illustrates the result of the global fit when considering such a configuration. As mentioned before, the only free parameters of the fit were the normalisation constants N_W and N_Z.

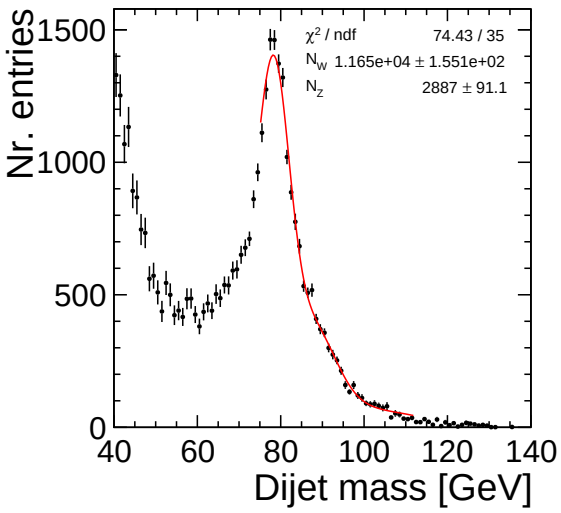


Figure 12.5: Result of the global fit, using the individual σ_W and σ_Z values $(\sigma_W^{res}, \sigma_Z^{res})$ obtained from the toy Monte Carlo study.

Furthermore, other options have also been considered: *(i)* the average of σ_W and σ_Z has been employed $(\sigma_{Average}^{res})$, *(ii)* both resolution values have been fixed to σ_W $(\sigma_W^{res}, \sigma_W^{res})$ and *(iii)* both resolution values have been set to the value of σ_Z $(\sigma_Z^{res}, \sigma_Z^{res})$. Thus, all these resolution configurations have been tested in turn when performing the global fit and the results are presented in table 12.1.

The χ^2/ndf of the global fit as well as the different values of the precision achiev-

Sample	Statistical precision			
Res. config.	$(\sigma_W^{res}, \sigma_Z^{res})$	$(\sigma_{Average}^{res})$	$(\sigma_W^{res}, \sigma_W^{res})$	$(\sigma_Z^{res}, \sigma_Z^{res})$
χ^2/ndf	2.13	3.41	2.35	5.35
$\tilde{\chi}_1^{\pm}$	1.33%	1.31%	1.29%	1.34%
$\tilde{\chi}_2^0$	3.15%	3.79%	3.17%	4.79%

Table 12.1: Table illustrating the statistical precision of the one-dimensional fit method for determining the chargino and neutralino cross-sections. The global 1D fit described above has been performed for different combinations of the detector resolution $\sigma_{W/Z}^{res}$.

able with the one-dimensional fitting method are also quoted in Table 12.1. The best global fit performance is obtained for the resolution configuration $(\sigma_W^{res}, \sigma_Z^{res})$. The average statistical precision of the method is 1.32% for the chargino case and 3.72% for the neutralino sample.

The one-dimensional fit method presented above unfortunately features several shortcomings. The Standard Model polynomial fit was orginally chosen because it performed the best from all the tested functions, i.e., without an underlying physics model. Furthermore, more importantly, as can be seen from figure 12.5, even after performing the toy Monte Carlo parameter determination, the global fit does not describe the data accurately.

It is relevant to note that this method was performed and studied within this analysis as a matter of legacy. However, the shortcomings of the one-dimensional fitting method as well as its precision can be improved when considering a more involved fitting procedure as presented in the following section.

12.2 Two-Dimensional Boson Mass Fit

The observable of interest for the two-dimensional fit method is the di-jet mass *before* performing the kinematic fit, but after applying the preselection cuts.

The topology of the signal events consists of four jets, produced by the hadronic WW or ZZ decays, and a large amount of mising energy due to the escaping $\tilde{\chi}_1^0$. There are three distinct possibilities to combine the four hadronic jets into di-jets. Clearly, only one of the three possible configurations describes the real process, while the other two are mis-associations.

Two different strategies concerning the di-jet pairing have been considered in the present study, as described in the following.

Fit with all Possible Jet Pairings

In the first strategy, *all* three possible di-jet combinations in an event are used. This implies that each event contributes with three entries (i.e., one for each di-jet pairing configuration) to the di-jet mass distribution. Only those events that have passed the kinematic fit convergence test are considered. This ensures a certain quality level of the data sample.

Figure 12.6 shows the relevant distribution for this study: the two-dimensional

representation of one di-jet mass versus the other, when all three possible di-jet combinations are taken into account. The number of events is colour-coded and presented on the z-axis. The distribution comprises the Point 5 signal events plus the SUSY and the Standard Model background events remaining *after* applying all the selection cuts.

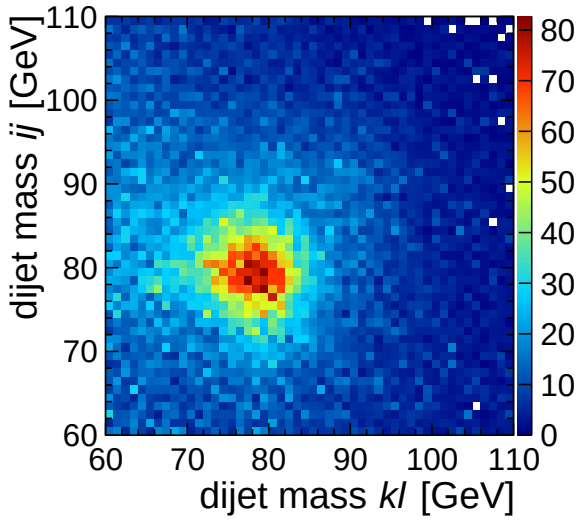


Figure 12.6: The total di-jet distribution of Standard Model background plus SUSY signal and SUSY background events obtained with the DBD version of the simulation and reconstruction software.

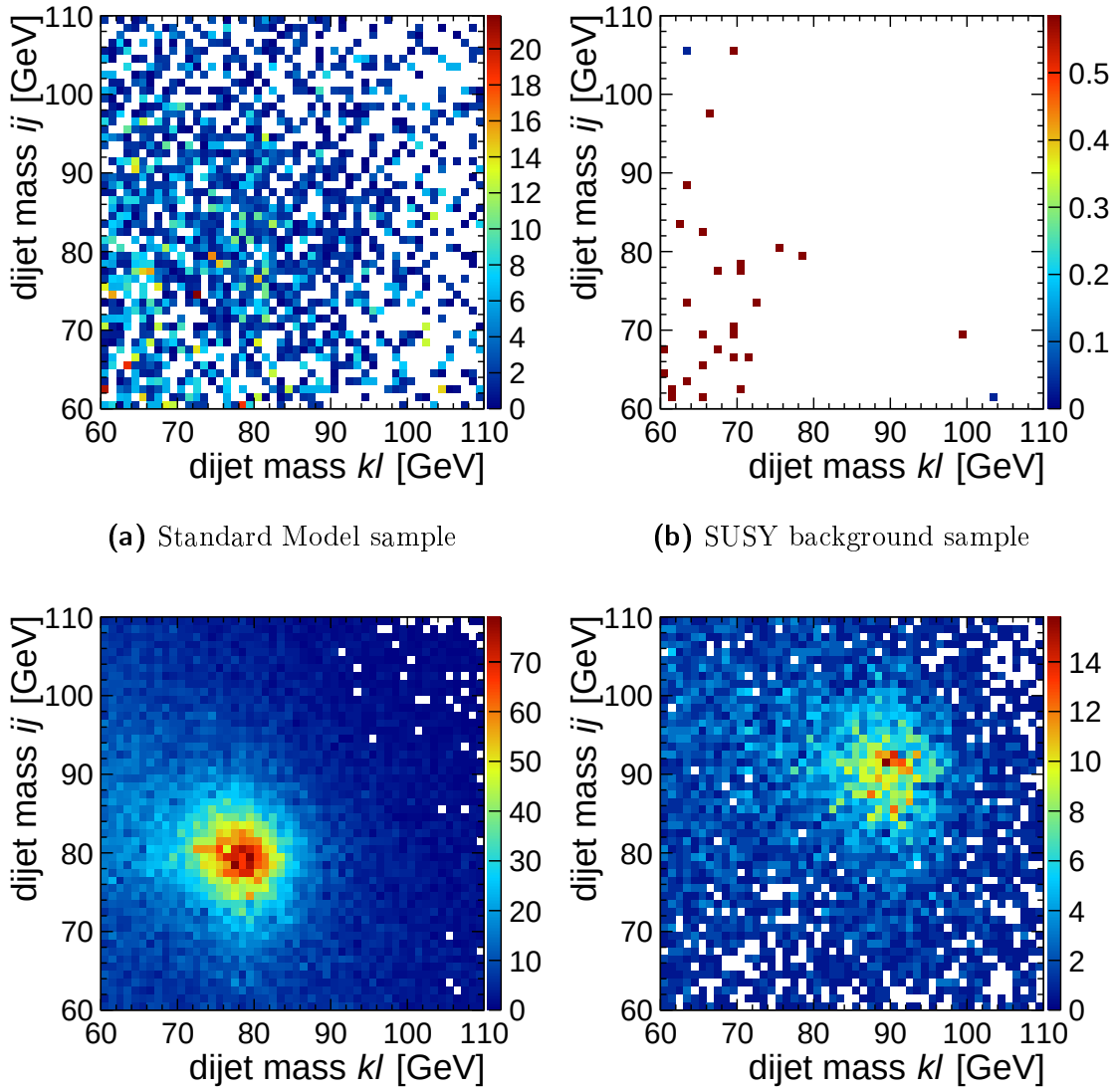
Each individual contribution to the total histogram shown in figure 12.6 is presented in figure 12.7: pannel (a) represents the Standard Model background that has survived the cuts, pannel (b) showcases the small SUSY background while pannels (c) and (d) show the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ contributions, respectively.

The cross-section is proportional to the number of events (equation 12.2). Consequently, the goal of this analysis is to determine the number of $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ events present in the total sample.

For this purpose, a *template* two-dimensional di-jet mass distribution can be produced from Monte Carlo data for each of the two ($\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$) signal samples. Based on these templates, a two-dimensional parametric function, $f_{2D}(x, y)$, can be formulated. The amount of $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ pair-like events present in the total distribution from figure 12.6 could be determined by fitting the newly obtained function to it.

Two assumptions were made when constructing the two-dimensional fitting function. The first assumption is that the Standard Model contribution is well known, perhaps from other sources, and thus it can be subtracted from the total (data) distribution before applying the fit.

The subtraction procedure is applied on a bin by bin basis. The total number of entries in any particular bin is: $N_{total} = N_{SM} + N_{signal}^{SUSY} + N_{bgrd}^{SUSY}$. The (expected) number of SM entries in that respective bin (N_{SM}) is subtracted from N_{total}. The statistical error is left unchanged, i.e. $Error_{bin_{x,y}} = \sqrt{N_{total}}$. This way, the fact



(a) Standard Model sample

(b) SUSY background sample

(c) Chargino sample

(d) Neutralino sample

Figure 12.7: Individual contributions to the total distribution in the di-jet mass plane, as depicted in figure 12.6.

that a statistical uncertainty in the SM estimation is still present no matter how accurately the SM data can be described is taken into account.

The second assumption is that the contribution from the SUSY background is very small or well reducible such that it can be safely neglected. Figure 12.7 (b), indicates that the SUSY background[1] is indeed negligible. Consequently, the fitting function can be formed with the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ templates alone:

$$f_{Fit}(x,y) = a \cdot f_{\tilde{\chi}_1^\pm}(x,y) + b \cdot f_{\tilde{\chi}_2^0}(x,y) \tag{12.4}$$

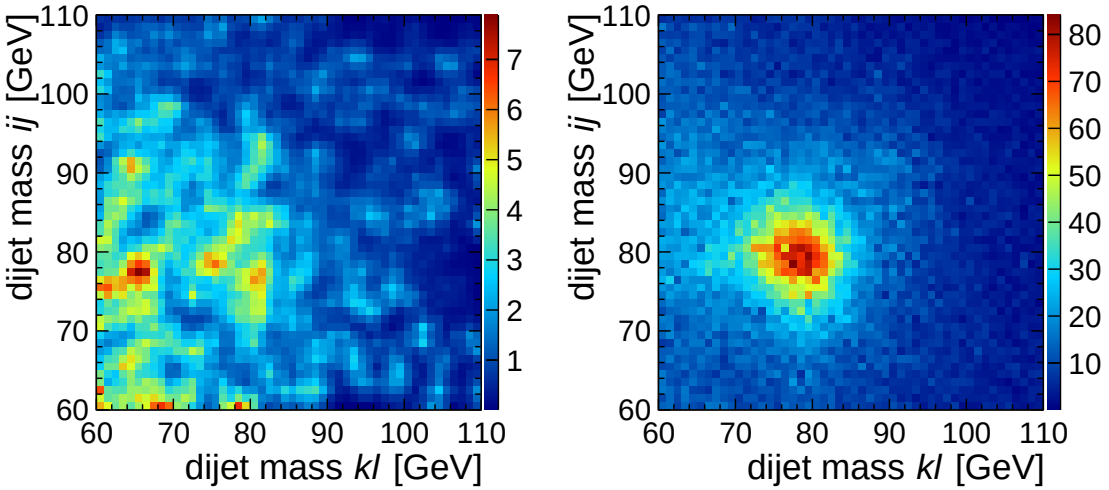
where a and b are the only free parameters. Variables a and b can be understood

[1] The fractional event numbers come from the previously discussed normalisation of all Monte Carlo data samples to the nominal ILC integrated luminosity and beam polarisation configuration (section 10.1).

as the fraction of the number of events from the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ template, respectively, that can be found in the total data distribution (for example, the one illustrated in figure 12.6). Thus, in an ideal case, when the number of events from the templates describe exactly the contributions in the total sample, the fit parameters a and b are expected to be equal to 1.

Once the template fitting function has been defined it can be applied to the data. However, due to the limited amounts of Monte Carlo data available, two issues must be taken into account.

First, the Standard Model background distribution shown in 12.7 (a) exhibits a non-negligible number of empty bins (marked in white). In principle, since the aforementioned distribution contains only those Standard Model events that have survived the SUSY selction cuts, it is expected that some bins will have zero or very few entries. However, due to the limited quantity of Standard Model Monte Carlo samples available, it is not obvious whether this effect is a result of applying the selection cuts or whether it is caused by low statistics. Clearly, this feature could be overcome in the future by producing more SM Monte Carlo data samples. Currently, in order to take this effect into account and level out possible statistical fluctuations, the Standard Model distribution is smoothed out. The smoothing procedure is analogous to the techniques applied in image processing with the purpose of blurring contours [238]. Thus, a Gaussian-like kernel expressed as a 5×5 matrix is convoluted with the original two-dimensional histogram over the entire histogram range. The kernel is acting as a low pass filter, thus evening out all high statistical fluctuations (e.g. bins with 10 entries or more completely surrounded by empty bins). Figure 12.8, (a) shows the result of the smoothing procedure. The total distribution from 12.6 is remade, using the smoothed SM distribution instead of the initial one, and the result is shown in 12.8 (b).



(a) Smoothed Standard Model distribution

(b) Total data distribution using the smoothed SM sample

Figure 12.8: The result of the Standard Model background smoothing procedure.

Second, due to the limited amount of Monte Carlo data, the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ di-jet distributions that are used as templates in the fitting function are the same as the signal contributions in the total distribution shown in figure 12.6. However, the templates and the data which is to be fitted with the template function must be statistically independent. In order to achieve statistical independence, two procedures have been considered. In both cases, a toy Monte Carlo study was considered for estimating the statistical precision of the two-dimensional template method: $\Delta\sigma = \frac{\sqrt{N_{data}}}{\epsilon \int \mathcal{L}dt}$.

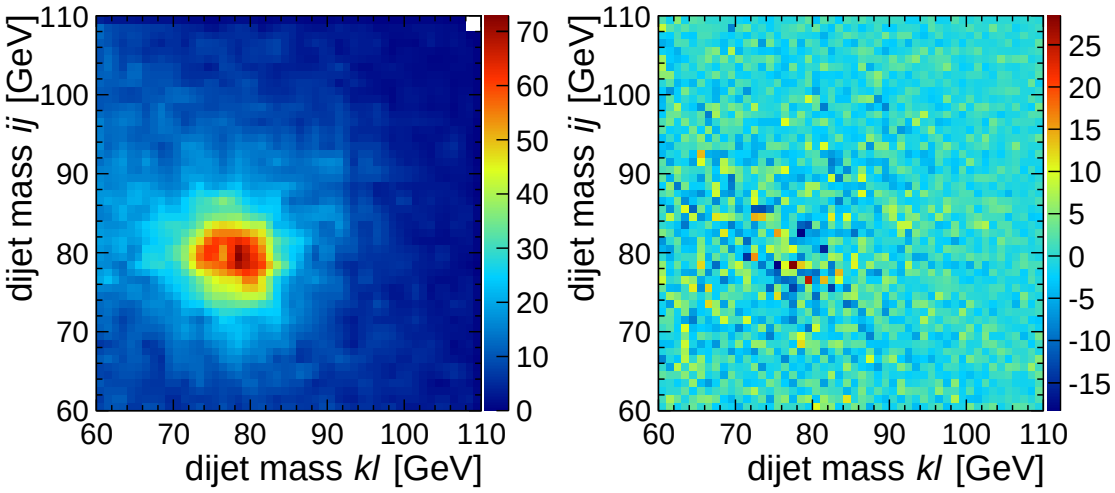
The first procedure consists of separating all available Monte Carlo samples in two: one part being treated as *data* while the other half is used as *template* for the fitting function. Two issues arose when this strategy was applied: *(i)* by dividing the already low amount of Monte Carlo the number of empty bins (or bins with very low numbers of entries) increased, *(ii)* the obtained mean event fractions (a and b) were shifted by $\approx 0.7\%$ with respect to the non-split case. This is most likely a statistical effect introduced by the sample splitting.

In order to avoid these issues, the second procedure relies on performing the toy Monte Carlo study with the whole data sample that survives the preselection cuts. The total data sample consisted of the smoothed SM background, illustrated in figure 12.8, pannel **(a)**, plus all the other contributions. It was then used as a basis to randomly generate 10^4 new two-dimensional di-jet mass distributions.

At first, the number of weighted events in the histogram ($N_{initial} = 32861.9$) was calculated. For each one of the new 10^4 distributions a new number of entries N_i^{new} was generated by randomly sampling a Poisson distribution with its mean equal to $N_{initial}$. In the next step, the initial histogram was randomised on a bin by bin basis. A Poisson distribution was sampled for each bin of the original histogram, where the total (initial) number of entries in that respective bin was considered as the Poisson mean. The current total number of entries in the new histogram is constantly compared to N_i^{new} and adjusted at each step such that the final number of events in the new histogram is exactly the desired N_i^{new}. The random generator that was used is RANLUX [239, 240] with luxury level 4, i.e. the highest level of luxury available.

For each newly generated two-dimensional di-jet distribution, the SM background was subtracted (while leaving the bin statistical errors unchanged) and the template fit was performed. Figure 12.9 illustrates in pannel (a), as an example, the fit result for one of the 10^4 performed fits. The fit residuals are shown in pannel (b) of the same figure. It is relevant to note that the residuals' distribution is randomly uniform and as such no distinctive pattern can be seen. This indicates that the fit performs well. The average value of the ratio of the fit χ^2 to the number of degrees of freedom is 1.229.

After the template fit is performed, the fit parameters a and b can be interpreted in terms of the corresponding cross-section. For illustration, let a be the fraction of events from the $\tilde{\chi}_1^{\pm}$ template that were "matched" in the total distribution. Since the cross-section is directly proportional to the number of events, the information on the cross-section that was considered when producing the $\tilde{\chi}_1^{\pm}$ template can be used to obtain the value of the cross-section in the total data



(a) Fit result **(b)** Residuals

Figure 12.9: The outcome of the two-dimensional template fit for one of the 10000 instances of the toy Monte Carlo study.

sample. The template cross-section is essentially scaled with the value of a, in this example.

This procedure is performed 10^4 times, for each randomly produced total distribution. Thus, 10^4 different cross-section values are obtained, as shown in figure 12.10. The average value determined for the $\tilde{\chi}_1^{\pm}$ fit parameter is $a = 1.001 \pm 0.009$ while for the $\tilde{\chi}_2^0$ case, the average value is $b = 1.004 \pm 0.029$.

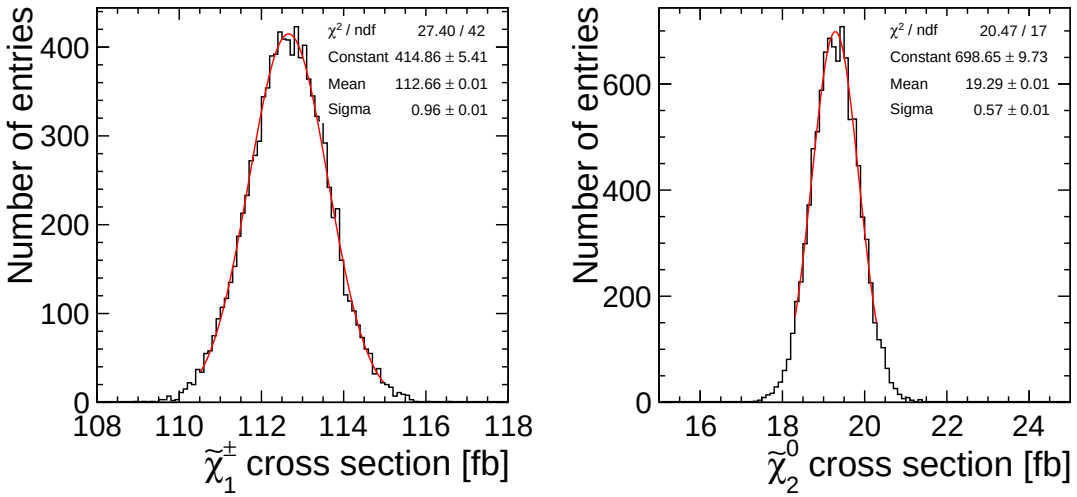


Figure 12.10: Result of the toy Monte Carlo study showing the obtained distributions of cross-section values.

The cross-section values obtained from the toy Monte Carlo form a Gaussian-like distribution, both in case of the $\tilde{\chi}_1^{\pm}$ as well as for the $\tilde{\chi}_2^0$. In order to estimate the statistical precision on the cross-section values, each of the two distributions can be fitted with a Gaussian function, where the width of the Gaussian is a

measure for the statistical precision. Thus, for the $\tilde{\chi}_1^\pm$ case, the cross-section can be measured with a statistical precision of:

$$\begin{aligned} \delta\sigma(e^+e^- \to \tilde{\chi}_1^+\tilde{\chi}_1^- \to W^+\tilde{\chi}_1^0 W^+\tilde{\chi}_1^0) &= 0.967\,\text{fb (precision equivalent of 0.85\%)} \\ \delta\sigma(e^+e^- \to \tilde{\chi}_2^0\tilde{\chi}_2^0 \to Z^0\tilde{\chi}_1^0 Z^0\tilde{\chi}_1^0) &= 0.573\,\text{fb (precision equivalent of 2.9\%)} \end{aligned}$$

These values are compatible with the statistical precision achieved in the original LOI study [109, 196]: 0.84% for the $\delta\sigma(e^+e^- \to \tilde{\chi}_1^+\tilde{\chi}_1^- \to W^+\tilde{\chi}_1^0 W^+\tilde{\chi}_1^0)$ case and 2.75% for $\delta\sigma(e^+e^- \to \tilde{\chi}_2^0\tilde{\chi}_2^0 \to Z^0\tilde{\chi}_1^0 Z^0\tilde{\chi}_1^0)$.

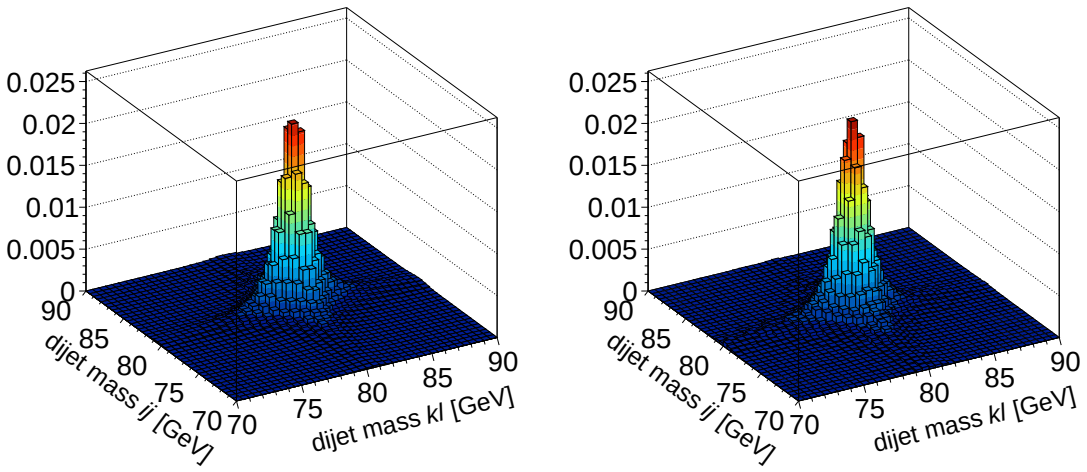
Lastly, it is important to note that the mean of the two Gaussian functions are very close to the model values for the cross-section, as presented in section 3.1: $\sigma(e^-e^+ \to \tilde{\chi}_1^-\tilde{\chi}_1^+) \times B.R.(\tilde{\chi}_1^-\tilde{\chi}_1^+ \to W^-\tilde{\chi}_1^0 W^+\tilde{\chi}_1^0) = 112.658\,\text{fb}$ (model value: 112.536 fb) while $\sigma(e^-e^+ \to \tilde{\chi}_2^0\tilde{\chi}_2^0) \times B.R.(\tilde{\chi}_2^0\tilde{\chi}_2^0 \to Z^0\tilde{\chi}_1^0 Z^+\tilde{\chi}_1^0) = 19.279\,\text{fb}$ (model value 19.243 fb). This comparison represents a useful cross check and the good agreement between the model and toy Monte Carlo obtained values shows that template fitting procedure performs well.

Fit with Best Jet Pairing

The *second strategy* for this study was to make use of the kinematic fit information in order to first determine the right di-jet configuration - i.e., the one which provides the best fit probability - and form the two-dimensional di-jet distributions considering only that particular di-jet configuration. The relevant di-jet mass distributions are the ones before applying the kinematic fit. The identical procedure for the template fit, as described above, is applied to these new distribution. The toy Monte Carlo study is repeated for 10000 new randomly generated samples.

The statistical error slightly increases when taking only the best di-jet configuration into account. In the $\tilde{\chi}_1^\pm$ case, the statistical precision deteriorates by approximately 15% ($\sigma_{best\ config.}(\tilde{\chi}_1^\pm) = 112.952 \pm 1.112\,\text{fb}$), while for $\tilde{\chi}_2^0$ it becomes worse by 3% ($\sigma_{best\ config.}(\tilde{\chi}_2^0\tilde{\chi}_2^0) = 19.670 \pm 0.587\,\text{fb}$). This observation can be explained by the fact that, despite its good performance, the kinematic fit cannot always indicate the correct di-jet association. In view of that, it is clear that in such cases one of the di-jet configurations that have a lower fit probability (and are discarded in the second strategy) is the correct one. This situation is accounted for in the first strategy discussed above, which in turn leads to a better statistical precision in that case.

It is also important to note that the construction of the fitting function and, implicitly, the shape of the templates are independent from the assumed supersymmetry scenario. A key model feature is, clearly, the mass difference (ΔM) between the almost mass degenerate $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ and the LSP. However, as long as the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ decays to on shell W and Z bosons are allowed, the di-jet mass distributions on which this study is based are not influenced by any other model assumptions. This is shown, as an example, in figure 12.11 where the relevant two-dimensional di-jet mass distribution from a Standard Model WW sample is compared on generator level to the full $\tilde{\chi}_1^\pm$ signal sample. Leaving aside the shift in the W boson mass featured in the $\tilde{\chi}_1^\pm$ sample (explained in section 10.1), the



(a) SM WW → qqqq **(b)** Chargino

Figure 12.11: Comparison of the di-jet mass distributions obtained from Standard Model $WW \to qqqq$ and Point 5 $\tilde{\chi}_1^+\tilde{\chi}_1^- \to qqqq\tilde{\chi}_1^0\tilde{\chi}_1^0$ events.

two distributions look very similar.

On the reconstruction level, the di-jet mass distributions are influenced by the mass difference, ΔM, in that it determines the amount of energy available to produce the WW/ZZ pairs. Depending on the available energy, the jets associated to the W/Z boson pairs can be more or less boosted in the event. In the SUSY model considered in this analysis, the W^+W^- and ZZ bosons are produced nearly at rest and thus rather central. Furthermore, the lower the boost the higher the probability to wrongly pair the jets becomes. This, in turn, translates to a higher imprecision, i.e., a worse resolution for the di-jet mass and, implicitly, for the gaugino mass determination. The mass resolution would only improve for larger mass differences ΔM. However, for the present analysis the gaugino masses are obtained, as shown in chapter 11, from a completely independent method which relies only on the value of the edge positions from the di-jet energy spectrum, without taking any assumptions on the cross-section into account. Therefore, in this context, it is considered safe and appropriate to apply the template method described above.

12.3 Systematic Uncertainties

As can be seen from equation 12.2, there are several sources for systematic uncertainties that must be taken into account:

$$\delta\sigma^2 = \left(\frac{\partial\sigma}{\partial N_{evts}}\right)^2 \delta N_{evts}^2 + \left(\frac{\partial\sigma}{\partial\epsilon}\right)^2 \delta\epsilon + \left(\frac{\partial\sigma}{\partial\mathcal{L}}\right)^2 \delta\mathcal{L} \qquad (12.5)$$

where δN_{evts} represents the uncertainties that reflect on the number of events, $\delta\epsilon$ denotes the (reconstruction and selection) efficiency and $\delta\mathcal{L}$ is the uncertainty of the luminosity measurement.

The systematic uncertainties that can affect the precision of the chargino and neutralino cross-section measurements via the number of counted events mainly stem from the uncertainties on the beam polarisation, on the beam energy and on the luminosity measurement.

At the ILC, the systematic uncertainty of the polarisation measurement is expected to be $\delta P/P \leq 0.25\%$ while the statistical errors are negligible [123].

The energy of the incoming beams is crucial for measuring $\sqrt{s}$ and essentially determines the whole energy scale of the collision events. The beam energy delineates the phase space available for the chargino and neutralino pair production to occur. At the ILC, a system of two energy spectrometers provides an accuracy on the beam energy measurement of $\delta E_{beam}/E_{beam} = 10^{-4}$ on a bunch-to-bunch basis.

The number of observed events is also directly related to the achieved luminosity. The luminosity at the ILC can be measured with a precision of $\delta\mathcal{L}/\mathcal{L} = 10^{-4}$.

The main source of systematic uncertainties, however, is expected to come from $\delta\epsilon$ which is comprised of the uncertainties on the reconstruction and selection efficiencies for the two signal channels considered in this analysis. In this context, the key reconstruction uncertainties could be evaluated by studying pure Standard Model $e^+e^- \to W^+W^- \to q\bar{q}q\bar{q}$ and $e^+e^- \to Z^0Z^0 \to q\bar{q}q\bar{q}$ samples.

For the integrated luminosity, polarisation and centre-of-mass values assumed in this analysis, the polarised cross-section for $W^+W^- \to q\bar{q}q\bar{q}$ production is $\approx$4.5 pb ($2.3{\cdot}10^6$ events) while for the $Z^0Z^0 \to q\bar{q}q\bar{q}$ case the cross-section is $\approx$0.4 pb ($2{\cdot}10^5$ events) [70].

A suitable control sample could be obtained by applying the selection procedure dedicated to the mass determination, described in section 10.4, while inverting the cut on the visible energy $100\,\text{GeV} < E_{visible} < 300\,\text{GeV}$ such that $E_{visible} > 300\,\text{GeV}$. This ensures that no SUSY events will be found in the SM control sample and that it is almost exclusively populated with four-fermion events. The statistical uncertainty in comparing the measured and the Monte Carlo data is 0.07% in the case of the $W^+W^- \to q\bar{q}q\bar{q}$ process and 0.2% for the $Z^0Z^0 \to q\bar{q}q\bar{q}$ process.

CHAPTER 13

Comparisons

The methods employed for the gaugino property determination, in the context of the Point 5 supersymmetry scenario, have been described in detail in the previous sections of this thesis. The techniques for measuring the production cross-sections (times the branching ratios) of the signal processes were presented in chapter 12. The determination of the three gauginos' masses was described in chapter 11. The analysis methods were exemplified exclusively using the DBD data set.

The same analysis techniques were applied on the $\mathsf{SGV}_{\mathsf{PFL}}$ and LOI data sets. For the former, the neutral visible energy scaling procedure described in section 8.5 was applied before the data preparation. In case of the latter, the data samples employed for the present analysis are those initially produced for the original ILD *Letter of Intent* [109] study. The analysis was repeated on this data in view of the new improvements described in the previous sections.

There are a few differences concerning the analysis strategy adopted for each of the three data sets. The application of slightly different strategies for the data preparation is motivated by the inherent differences between the data samples, overviewed in table 10.1. A summary of the data preparation and analysis stages is presented in table 13.1 and the relevant differences are indicated, where necessary.

13.1 Event Selection

The observables employed for the event selection as well as the corresponding signal regions were presented in detail in section 10.4.

There are four extra cuts applied in the case of the DBD and $\mathsf{SGV}_{\mathsf{PFL}}$ data samples: *(i)* the requirement for no isolated leptons to be found in the event, *(ii)* the limits on the number of reconstructed particles (PFOs) found in the event, *(iii)* the cut on the number of tracks with transverse momentum $p_T < 1\,\mathrm{GeV}$ and *(iv)* the upper limit on the thrust value in the event. Their main purpose was to refine the event selection for the DBD analysis and help further remove

Analysis Stage		LOI	DBD	SGV$_{PFL}$
Data Preparation	Remove $\gamma\gamma \to$ *hadrons* bgrd.	-	Yes	-
	Cluster events in 4 jets	Yes	Yes	Yes
	Run kinematic fit ($M_{j_1j_2} \equiv M_{j_3j_4}$)	Yes	Yes	Yes
	Perform event selection	Yes	Yes	Yes
	Extended selection: 4 extra cuts	-	Yes	Yes
Cross-section Measurement	No further signal separation	Yes	Yes	Yes
	Observable: di-jet mass before kin. fit (all 3 combin.)	Yes	Yes	Yes
Mass Determination	Signal separation using χ^2 cut (calc. w. di-jet mass *before* kin. fit)	Yes	Yes	Yes
	Observable: di-jet energy *after* kin. fit	Yes	Yes	Yes

Table 13.1: Overview of the data preparation and analysis steps carried out in the Point 5 study. Wherever necessary, the differences in the treatment of the three data sets are indicated.

the $\gamma\gamma \to$ *hadrons* background. This background is not present in the SGV data, however the extra cuts were still applied for convenience and comparison purposes. For backwards compatibility reasons these four extra cuts were not applied on the LOI data set.

The selection cuts and the effects of their application, in terms of numbers of surviving events, were illustrated using the DBD data set in the cutflow table 10.6 presented in section 10.4. The same event selection was also performed on the SGV$_{PFL}$ data set. The original selection procedure was repeated in the LOI case. The results were then compared and the most prominent difference observed concerns the visible energy observable.

The Visible Energy

Comparing the amounts by which the numbers of Standard Model background events decrease as the selection cuts are progressively applied, it was observed that almost 50% more LOI events survive the cut on the visible energy (100 GeV $< E_{vis} <$ 300 GeV) in comparison to the DBD and SGV$_{PFL}$ cases, for the following final states: $W^+W^- \to qqqq$, $Z^0Z^0 \to qqqq$ and $W^\pm Z^0 \to qqqq$.

For all three final states, the visible energy signal region corresponds to the tail of the observable's distribution, as shown in figure 13.1.

Firstly, it can be seen from figure 13.1, that the peaks of the three LOI, DBD and SGV$_{PFL}$ distributions are well compatible. However, there is a significant discrepancy between the three data sets in the low energy (tail) region.

The visible energy is clearly highly sensitive to the $\gamma\gamma \to$ *hadrons* background removal, especially in this area of the distribution. The parameters of the exclusive k_T algorithm used for this purpose were optimised for obtaining the best results in terms of the di-jet energy and mass distributions (figure 10.2). The

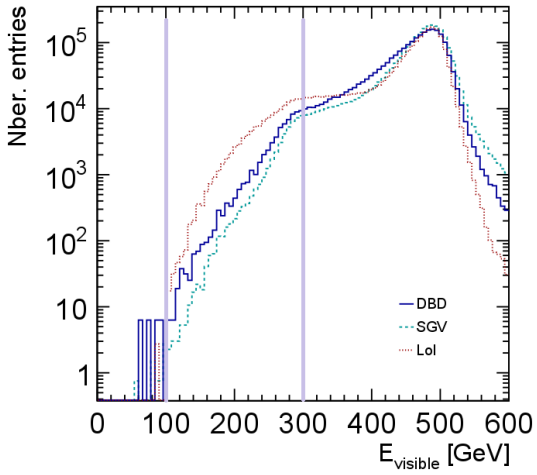


Figure 13.1: Comparison of the visible energy distributions obtained from the LOI, DBD and SGV_{PFL} Standard Model $W^+W^- \to qqqq$ data, after the preparation stage. The events found within the region marked by the vertical lines (100-300 GeV) are kept (survive the cut), while the rest are discarded.

tail of the visible energy distribution was not considered in the optimisation process. Therefore, in case of the DBD data set, it is quite likely that the observed discrepancy is an effect of the $\gamma\gamma \to hadrons$ background removal procedure.

In contrast, the SGV_{PFL} data does not contain the $\gamma\gamma \to hadrons$ background overlay. Nevertheless, the visible energy is also very sensitive to detector effects like, e.g., coverage close to the beam pipe, non-instrumented areas where neutral particles might escape detection, etc. It is possible that the differences between the fast and the full simulation distributions, especially the ones observed in this region, are a reflection of the different ways in which these detector geometry aspects are implemented in the three simulation programs.

Selection Efficiency and Sample Purity

The outcome of the event selection was evaluated in terms of the efficiency, ϵ, and purity, π, according to the definitions given in section 10.4.

As indicated in chapters 12 and 11, the analysis strategy applied in the Point 5 study considers two stages in the event selection. The first twelve (or eight, in case of the LOI data) cuts constitute the first level, i.e., "Level 1", and they are used in selecting events for the mass determination. In the second stage, "Level 2", *all* selection cuts are applied and the surviving events are input to the cross-section fit.

The efficiency and the percentage remaining in the sample were calculated for all three data sets, LOI, DBD and SGV_{PFL}, at both stages of the selection. The results are summarised in table 13.2.

It can be seen from table 13.2 that the application of four more additional cuts on the DBD and SGV_{PFL} data does not decrease the selection efficiency of the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ signal samples. In fact, it is up to 3% better than in the LOI case.

Furthermore, the differences observed in the outcome of the visible energy

Selection stage			$\tilde{\chi}_1^{\pm}$-had.	$\tilde{\chi}_2^0$-had.	SUSY other	SM bgrd.
Level 1	Efficiency [%]	LOI	88	91	9	0.5
		DBD	91	91	11	0.5
		SGV$_{\text{PFL}}$	90	90	11	0.4
	% final sample	LOI	12	2	3	82
		DBD	15	3	5	78
		SGV$_{\text{PFL}}$	18	3	5	74
Level 2	Efficiency [%]	LOI	70	74	7	0.1
		DBD	72	73	8	0.1
		SGV$_{\text{PFL}}$	72	72	8	0.1
	% final sample	LOI	26	5	6	62
		DBD	27	5	8	59
		SGV$_{\text{PFL}}$	30	5	8	57

Table 13.2: The efficiency and purity values calculated for each of the three considered data sets, at the two relevant stages in the event selection.

selection cut, described above, manifest as an up to 30% lower purity of the signal samples in the LOI case.

13.2 Mass Measurements

The Monte Carlo data used for the mass determination must pass the first stage (Level 1) of the event selection. Furthermore, since in the Point 5 scenario the chargino and neutralino pair production are both treated as signal, a further separation of the data samples, i.e., into $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ candidates, is required.

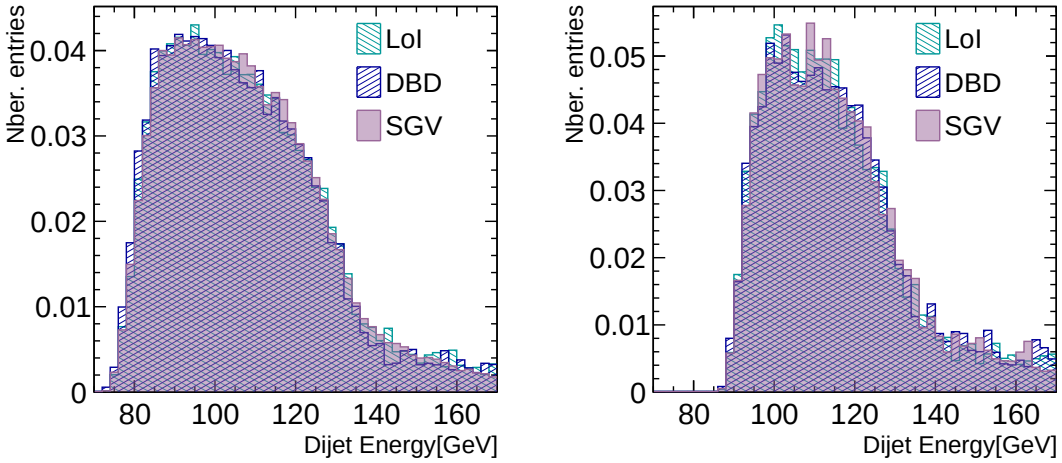
The separation procedure was described in section 11.2 and it involves the calculation of two χ^2 variables which are then used to evaluate whether the di-jet masses reconstructed in an event are more compatible with the W boson mass value or, on the contrary, with the Z boson mass. The same method was used for all three data sets considered in the comparison. The efficiency and purity obtained were presented for the DBD case in table 11.1. The values for the LOI and SGV$_{\text{PFL}}$ data sets are compatible, as shown in table 13.3.

The relevant observable for determining the gaugino masses is the reconstructed di-jet energy. The masses of the $\tilde{\chi}_1^{\pm}$, $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ can be calculated with the end-point method described in section 11.1, once the positions of the edges of the chargino and neutralino energy spectra are known.

The di-jet energy distributions obtained from the LOI, DBD and SGV$_{\text{PFL}}$ data sets are compared in figure 13.2. The most remarkable outcome of the comparison is the strong similarity between the three distributions, exhibited both in the $\tilde{\chi}_1^{\pm}$ and in the $\tilde{\chi}_2^0$ selections. This is notable in the context of the significant

	Simulation	$\tilde{\chi}_1^{\pm}$-had.	$\tilde{\chi}_2^0$-had.
Efficiency [%]	LOI	48	39
	DBD	45	39
	SGV_{PFL}	44	37
Purity [%]	LOI	71	60
	DBD	70	64
	SGV_{PFL}	76	67

Table 13.3: The efficiency and purity values of $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ in the candidate sample, after the signal separation.



$\tilde{\chi}_1^{\pm}$ Selection $\tilde{\chi}_2^0$ Selection

Figure 13.2: Comparison of the di-jet energy spectra obtained from the LOI, DBD and SGV_{PFL} data sets. The shown distributions were obtained after the standard data preparation and χ^2-based sample separation. All distributions were normalised to an integral of one.

discrepancy between the jet energy resolution provided by the PandoraPFA [7] implementations in the two full simulation versions, `LoI` and `DBD`, and the more pessimistic performance provided by $\texttt{SGV}_{\texttt{PFL}}$ (section 8.6). It was assumed that the ≈50% worse (on average) jet energy resolution obtained with $\texttt{SGV}_{\texttt{PFL}}$ would have a more sizeable impact on the edges of the energy spectra.

The new and more stable FIR-filter method, presented in section 11.4, was applied on all three data sets to extract the spectrum edge positions. The results are presented in table 13.4. The good compatibility between the di-jet energy distributions shown in figure 13.2 is also reflected in the determined edge positions.

It can be seen that the edge values obtained for all three data sets under comparison are well compatible within their respective uncertainties. This provides a quantitative confirmation of the strong resemblace between the di-jet energy distributions observed in figure 13.2.

Data	$\tilde{\chi}_1^{\pm}$		$\tilde{\chi}_2^0$	
	Low Edge (GeV)	High Edge (GeV)	Low Edge (GeV)	High Edge (GeV)
LOI	80.4 ± 0.2	129.9 ± 0.7	92.3 ± 0.4	128.3 ± 0.9
DBD	79.5 ± 0.2	129.5 ± 0.7	92.1 ± 0.3	128.4 ± 0.8
SGV_{PFL}	80.4 ± 0.2	128.6 ± 0.9	92.4 ± 0.3	126.9 ± 1.3

Table 13.4: The values of the edge positions determined by applying the FIR-filter technique on the LOI, DBD and SGV_{PFL} di-jet energy distributions.

For any of the four dijet-energy spectrum edges, its position and the precision with which it can be determined are affected by several factors. As mentioned in chapter 11, there are unavoidable smearing effects coming from the initial state radiation (ISR), the beam energy spectrum and the gauge boson natural width which are present in all three data sets. It must be noted that, while the LOI and DBD Monte Carlo data were produced with two slightly different beam energy spectra (section 6.2), the SGV_{PFL} data was produced with the same beam spectrum as the DBD.

The jet energy resolution, as defined in equation 4.2 (and discussed in section 4.1.2) was also expected to have an impact. However, as shown in table 13.4, the edge values obtained for the three different data sets are well compatible with each other. The same jet clustering algorithm was used on all three data sets. Furthermore, the association of the individual jets in pairs was performed identically, based on the same criterion. Thus, the only difference between the LOI, DBD and SGV_{PFL} data sets, in terms of the jet energy resolution, is the significantly different Particle Flow performance provided by the fast simulation SGV_{PFL} with respect to the two full simulation versions. Nevertheless, this is in fact not perceptible at the analysis level. Consequetly, it can be concluded that, for this particular physics scenario, the jet energy resolution is most probably dominated by jet clustering effects and the inevitable uncertainties in the jet pairing. The effects of the significantly worse Particle Flow performance provided by the fast simulation SGV_{PFL} are obscured by the more prominent aforementioned contributions.

Simulation	$\tilde{\chi}_1^{\pm}$ Mass (GeV)	$\tilde{\chi}_2^0$ Mass (GeV)	$\tilde{\chi}_1^0$ Mass (GeV)
Model	216.5	216.7	115.7
LOI	218.5 ± 3.6	220.5 ± 1.5	118.1 ± 1.1
DBD	218.4 ± 3.2	221.1 ± 1.3	118.3 ± 0.9
SGV_{PFL}	218.2 ± 3.9	220.5 ± 1.3	118.8 ± 1.3

Table 13.5: Comparison of the gaugino mass values calculated for the three data sets: LOI, DBD and SGV_{PFL}. The Point 5 model values (chapter 9) are also shown for reference.

Lastly, it must be noted that the application of the kinematic fit in the Point 5 analysis (section 10.3) also reduces the extent with which the different Particle Flow performances affect the di-jet observables.

For completion, the gaugino mass values were computed using the extracted edge positions presented in table 13.4. The shown statistical uncertainty was determined from toy Monte Carlo studies, as described in chapter 11. The results are presented in table 13.5. As expected, it can be observed that the comparison of the calculated mass values determined for the three data sets reveals that they are also well compatible within their statistical uncertainties.

13.3 Cross Section Determination

The reconstructed di-jet mass is the observable used in the determination of the cross-sections[1] for the two Point 5 signal processes. The goal is to determine the number of events from the total data sample that can be classified as $\tilde{\chi}_1^{\pm}$-like and $\tilde{\chi}_2^0$-like, respectively.

For this purpose, a template fit, described in section 12.2, is applied to the two-dimensional distributions formed by presenting one reconstructed di-jet mass as a function of the other. The two dimensional di-jet mass distributions obtained from the LOI, DBD and SGV_{PFL} data are illustrated in figure 13.3.

All selection cuts (Level 2) are applied on the data samples used for the cross-section determination. Furthemore, as mentioned in section 12.2, all three possible jet pairing combinations are used in the distributions. This implies that each event contributes with three entries in the final histogram.

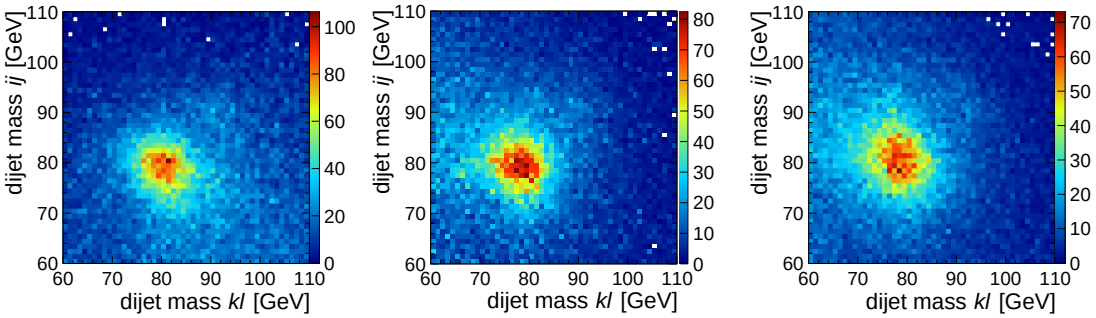


Figure 13.3: Two dimensional di-jet mass distributions for the three data sets considered for comparison. They constitute the input to the two-dimensional template fit.

The two free parameters of the template fit applied on the two-dimensional distributions shown in figure 13.3 are a measure for the amount of $\tilde{\chi}_1^{\pm}$-like and $\tilde{\chi}_2^0$-like events that were found. Based on these numbers, the corresponding cross-section values can be determined.

In a toy Monte Carlo study, presented in detail in section 12.2, the template fit is applied 10000 on randomly generated new di-jet mass distributions. The study was performed in the same conditions for all three data sets considered for comparison. The results are presented in figure 13.4 and in table 13.6.

[1] In fact, the analysis measures the cross-section times the branching ratio. However, for convenience, this is referred to simply with the term "cross-section".

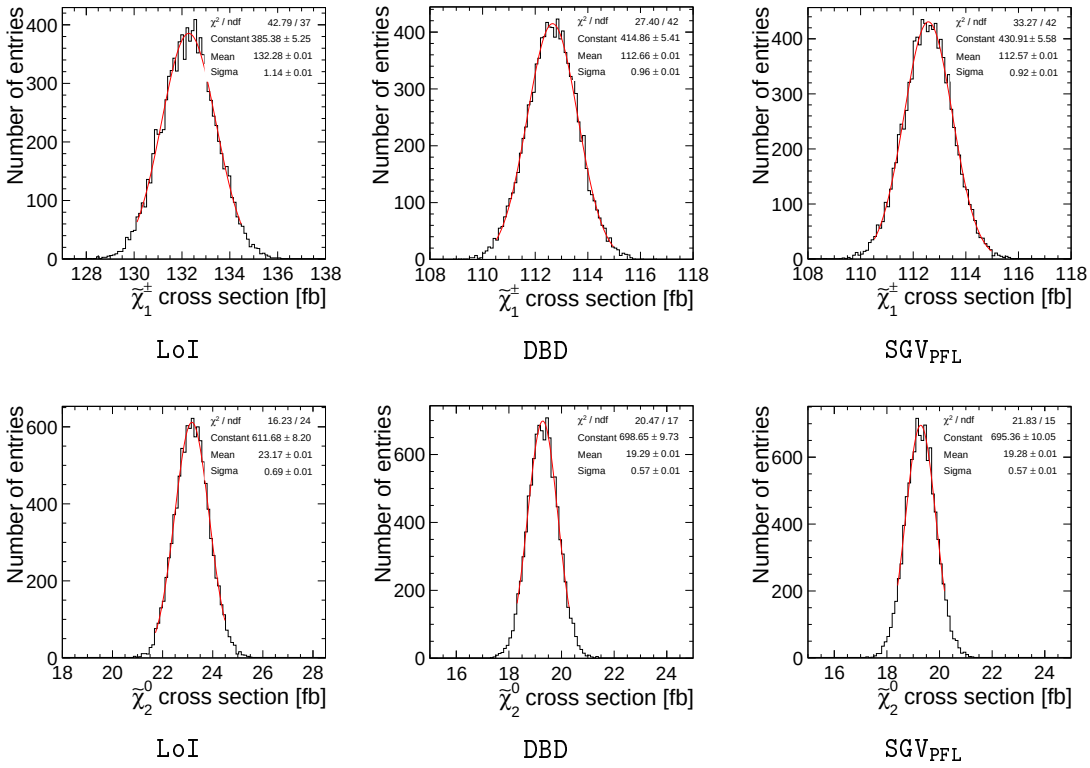


Figure 13.4: Comparison of the results obtained from the toy Monte Carlo study performed on the LOI, DBD and SGV$_{\mathrm{PFL}}$ two-dimensional di-jet mass distributions.

Simulation	$\tilde{\chi}_1^{\pm}$ cross-section (fb)	$\tilde{\chi}_2^0$ cross-section (fb)
Generator level	132.15	22.79
LOI	132.28 ± 1.14	23.17 ± 0.69
Generator level	112.54	19.2
DBD	112.66 ± 0.96	19.29 ± 0.57
SGV$_{\mathrm{PFL}}$	112.57 ± 0.92	19.28 ± 0.57

Table 13.6: The cross-sections and statistical uncertainty values determined in a toy Monte Carlo study performed for all three data sets used in the comparison. The LOI numbers are different than the DBD and SGV$_{\mathrm{PFL}}$ ones due to the different beam energy spectrum used in the Monte Carlo data generation.

The comparison of the DBD and SGV$_{\mathrm{PFL}}$ results with the values obtained from the LOI data is made more difficult by the use of different beam energy spectra in the data production. Nevertheless, they can still be compared to each other.

The same remarkable compatibility as the one observed in the gaugino mass study presented above is found. The DBD and SGV$_{\mathrm{PFL}}$ cross-section values agree very well within their statistical uncertainties. This appears to confirm the previously reached conclusion. The significant discrepancy between the more pessimistic Particle Flow performance provided by the fast simulation SGV$_{\mathrm{PFL}}$ in comparison to the DBD version of the full simulation is not detectable at the analysis level in the Point 5 study.

CHAPTER 14

Conclusions and Outlook

Performing precision measurements of the relevant Standard Model observables at the International Linear Collider (ILC) and searching for any potential deviations from their predicted values could provide a gateway to physics beyond the Standard Model.

Many reactions that are interesting for physics studies have final states with a high multiplicity of hadronic jets. For the level of precision required at the ILC, the traditional approach to calorimetry and to measuring the hadronic jets' energies is no longer satisfactory. A potential solution, known as *Particle Flow* (chapter 4), emerged in the CERN Large Electron Positron (LEP) collider era. While various experiments like, e.g., ALEPH [103], ZEUS [104] or H1 [95] have employed Particle Flow techniques, the detector concepts proposed for future electron-positron linear colliders like ILC or CLIC [52] are the first to have been specifically optimised for Particle Flow reconstruction.

The factors that influence the precision with which the energy of a hadronic jet (i.e., the jet energy resolution) can be measured were discussed in section 4.1.2. The most prominent contributions come from the inevitable uncertainties that occur in the jet clustering stage and from the intrinsic resolution of the detector subcomponents where the energies of the particles that form the jet are measured. The Particle Flow concept addresses the latter and proposes that the jet energy resolution could be significantly improved if the energy of each reconstructed particle was determined based on the measurements performed in the sub-detector that provides the best (intrinsic) resolution for that specific particle type. The Particle Flow paradigm constituted the main focus of the work presented in this thesis.

The Particle Flow approach has important implications not only for the detector design, but also regarding the methods and algorithms used in the event reconstruction. The former is clearly demonstrated by the design concept of the International Large Detector [123] (described in chapter 5) which constituted the

experimental context of the studies performed for this thesis. The latter manifests in the form of the reconstruction software PandoraPFA [7] which represents the most evolved implementation of a Particle Flow algorithm.

The main issue that limits the Particle Flow approach is the unavoidable *confusion* that occurs when reconstructing high energy events with a large particle multiplicity. In such dense environments, the pattern recognition algorithms used in the Particle Flow reconstruction occasionally fail to resolve and correctly identify energy deposits produced in the detector by different particles. This inevitably occurs in PandoraPFA and it was shown that the confusion effects increase with the jet energy [7].

This thesis first investigated some aspects of the Particle Flow confusion in the context of the fast simulation program `SGV` [6] (chapter 7) which is used to continually update the physics potential achievable with the ILD experiment.

In its default implementation, `SGV` does not emulate the confusion effects that unavoidably occur in PandoraPFA. Firstly, the impact of this simplification on the jet energy resolution was investigated using the so-called $Z \to uds$ data sets. As discussed in section 6.6.2, using this type of events is advantageous since it allows all the unrelated contributions to the jet energy resolution to be safely neglected such that the Particle Flow performance alone can be studied. The jet energy resolution obtained with the fast simulation SGV was compared with the performance of two subsequent implementations of the ILD full simulation, `LoI` and `DBD`. The two full simulations employ two consecutive versions of the PandoraPFA reconstruction algorithm, described in chapter 6. It was found that, in its default running mode, `SGV` underestimated the confusion effects such that the jet energy resolution obtained with it is up to 30% better than the one attainable in full simulation (8.1).

Considering the precision envisioned at the ILC, this discrepancy had to be addressed. The level of realism provided by `SGV` had to be increased such that its output becomes compatible with the ILD full simulation. The first approach consisted of studying the confusion-related behaviour of PandoraPFA in full simulation with the aim of parametrising it. By implementing these parametrisations in the fast simulation it was expected that the confusion effects observed in the Particle Flow reconstruction could be emulated in `SGV`. A first proof-of-principle study was carried out for this thesis. It demonstrated that it is indeed feasible to formulate parametrisations based on observables that are also easily accessible in the fast simulation (section 8.3.1).

After a set of parametrisations were implemented in `SGV` [6], another Particle Flow comparison study was carried out in this thesis (section 8.4). It revealed that, as a result of the parametrisations: *(i)* the jet energy resolution obtained in `SGV` (with confusion) is on average 55% worse in comparison to the full simulation and *(ii)* the distribution of the total visible energy reconstructed in an event was shifted towards higher values in comparison to both the full simulation and the default `SGV` running mode. It was found that this shift towards higher visible energy values can be corrected and this thesis contributed further by developing an energy correction procedure described in section 8.5. A final Particle Flow

performance study compared the results of the two full simulation versions with the output of SGV with the confusion emulation and *after* the energy correction. It was concluded that while the shift of the visible energy distribution was indeed corrected, this comes at the cost of the jet energy resolution which is up to 9% worse after the energy correction.

It was decided that the significantly worse jet energy resolution provided by the fast simulation after the implementation of the confusion parametrisations could be considered a pessimistic case. Thus, it was determined that the Particle Flow performance of SGV should be further evaluated in comparison to the two full simulation versions in the context of a more involved physics scenario. For this purpose, the Point 5 supersymmetry model, described in chapter 9, was chosen.

The Point 5 scenario considers the chargino ($\chi_1^{\pm}$) and neutralino (χ_0^2) pair production at the ILC. The gauginos decay almost exclusively to the corresponding gauge bosons and the lightest supersymmetric particle (LSP). The hadronic decay modes of the two gauge bosons were specifically chosen as signal for the analysis in order to challenge the Particle Flow performance.

The determination of the gaugino masses and of the signal processes' cross-sections had already been performed previously in a study carried out for the ILD *Letter of Intent* document [109]. The original analysis methods were at first repeated in the context of this thesis. As they were applied on Monte Carlo data produced with the (more recent) DBD version of the full simulation, a series of improvements emerged.

Firstly, the event selection (chapter 10) was refined in order to remove the newly considered $\gamma\gamma \rightarrow$ *hadrons* background. Concerning the cross-sections' measurements (chapter 12), a new toy Monte Carlo approach was developed in order to evaluate the statistical uncertainties. In the original (*Letter of Intent*) study, the evaluation of the statistical uncertainty required a "splitting" of the Monte Carlo data due to the low amount of simulated data available. This could introduce an unwanted bias in the evaluation. This issue is circumvented with the new approach.

For the gaugino mass determination (chapter 11), measuring the positions of the edges of the di-jet energy spectra is crucial. It was demonstrated that the method used in the original *Letter of Intent* study, which relied on a one-dimensional fit applied to each energy spectrum, was too sensitive to small fluctuations in the energy distributions and, hence, proved unstable. A completely new method was developed (section 11.4) for this thesis. It involves the application of a finite impulse response (FIR) filter in order to determine the inflection points in the di-jet energy spectra.

In addition, an edge calibration procedure was performed (section 11.5). By calibrating the edges determined in the analysis, i.e., at reconstruction level, with respect to the ones obtained from purely kinematic calculations, the added impact on the edge position of the ISR emission, beam energy spectrum and simulation and reconstruction effects was estimated. It was found to be of the order of approximately 2%.

Furthermore, while the Point 5 analysis methods, applied and developed in this

thesis, are illustrated only in the DBD context, the same methods were applied on the data obtained with the LoI version and with SGV. The studies presented in this thesis (chapter 13) represent the first consistent direct comparison of the three simulations carried out on the analysis level.

It was concluded that, in case of the Point 5 scenario, the large discrepancy between the jet energy resolution achieved in SGV (after the confusion emulation implementation) in comparison to the full simulation performance is not perceptible on an analysis level. The impact of the significantly more pessimistic Particle Flow performance provided by SGV on the jet energy resolution is obscured by the much more prominent contributions from the jet clustering and jet-pairing effects.

It must be noted that the implementation of the Particle Flow confusion emulation in SGV studied in this thesis represents only the first stage of a work in progress. New analysis tools are presently being developed with the goal of improving the understanding of confusion as it occurs in PandoraPFA. For instance, the TrueJet software tool [91] could be employed to evaluate the effects of the jet clustering errors on the jet energy resolution. This could provide the means to make a more precise quantitative estimate of the contribution coming from the Particle Flow confusion since the size of the contribution stemming from the jet clustering effects would then be known.

The comparison studies presented in this thesis also demonstrated the need to compare the detector and the Particle Flow performances using a definition of the jet energy resolution that is more closely applicable on the analysis level than the standard one (equation 4.4), employed in the context of the $Z \to uds$ samples. This is particularly important in the context of the ongoing ILD optimisation studies.

In view of the interest announced by Japan to potentially host the ILC project, the detector design is presently optimised with the aim of finding the best achievable balance between its cost and the physics performance. The findings presented in this thesis, especially regarding the prominent contributions of the jet clustering effects to the total jet energy resolution, inform the choice of processes used to benchmark the ILD performance.

As the Particle Flow story presented in this thesis, from the concept's first principles to the determination of gaugino properties, draws to an end, the ILC story is just ready to start.

Acknowledgements

Lecţia despre cub
Se ia o bucată de piatră,
se ciopleşte cu o daltă de sînge,
se lustruieşte cu ochiul lui Homer,
se răzuieşte cu raze,
până cubul iese perfect.
După aceea se sărută de nenumărate ori cubul
cu gura ta, cu gura altora
şi mai ales cu gura infantei.
După aceea se ia un ciocan
şi brusc se fărîmă un colţ de-al cubului.
Toţi, dar absolut toţi zice-vor:
- Ce cub perfect ar fi fost acesta
de n-ar fi avut un colţ sfărîmat!
Nichita Stănescu,
"Operele Imperfecte", 1979

The Lesson About Cubes[a]
One takes a stone,
Carves it with a chisel of blood,
Polishes it with Homer's eyesight,
Etches it with sunlight,
Until one achieves a perfect cube.
Afterwards, one kisses the cube,
countless times,
With one's own mouth, with someone else's mouth,
But most importantly, with an innocent's mouth.
Finally, one takes a hammer,
And, suddenly, smashes a corner of the cube.
All, every single one of them, will say:
What a perfect cube this could have been,
But for the broken corner!
Nichita Stănescu,
"Imperfect Works", 1979

[a] Translated to English with the help of my dear friend, Dr. Amanda Steber.

Many people have accompanied me on my journey that led to the completion of this thesis. I would like to express my gratitude to each one of them in the following.

I thank *Dr. Ties Behnke* for giving me the opportunity to pursue my PhD thesis in the DESY FLC group.

I am very grateful to *Prof. Caren Hagner* and *Prof. Elisabetta Gallo* for kindly accepting to be part in my defence committee.

My deepest gratitude goes to my supervisor, *Dr. Jenny List*. I thank her for encouraging me to embark upon this journey, for her guidance and all the enlightening discussions we have had over the years. I have learned very much from

her, both scientifically and also personally. Most importantly, I thank her for her confidence and unvwavering support.

During the PhD students' mentoring programme I have received fruitful advice from my mentors: *Dr. Karsten Büßer*, *Dr. Felix Sefkow* and *Dr. Isabell Melzer-Pellmann*. I kindly thank them for the very instructive discussions.

Concerning the work presented in this thesis, I would like to thank:

- *Dr. Mikael Berggren* - for his thorough explanations and ample discussions both concerning physics topics in general and regarding the fast simulation `SGV`.
- *Dr. Katja Krüger* - for always helping me to learn more with clear answers to all my analysis and physics-related questions, for her helpful advice and support.
- *Prof. Gudrid Moortgat-Pick* - for always having an open door for my many theory-related questions and for her encouragement.
- *Stefano Caiazza* - for the many helpful discussions about FIR filters and their implementation.
- *Dr. Daniel Dercks* - for his invaluable help with the `CheckMATE` analysis.

I would also like to thank *Dr. Christoph Rosemann*, *Dr. Maria Aldaya Martin* and *Dr. Katarzyna Wichmann* for their support and encouragement.

I also thank *Dr. Annika Vauth* for her friendliness and for giving me the opportunity to take part in her polarimeter prototype test beam campaign at DESY, from which I have learned a lot.

I am very grateful to my office mates: *Moritz Habermehl*, *Suvi-Leena Lehtinen*, *Dr. Shin-ichi Kawada* and *Swathi Kollassery* for the great office atmoshpere, for their continuous support and understanding.

My gratitude and appreciation go to *Dr. Moritz Beckmann*, *Dr. Hale Sert* and *Dr. Benedikt Vormwald*. From good office mates, they have become three of my closest and dearest friends. I thank them for their wonderful sense of humour, for their help and support, for many great discussions about physics, life and the Universe. Most of all, I thank them for their friendship.

I am also very thankful to *Mrs. Andrea Schrader* for her kindness, unwavering friendship and support throughout all these years.

I thank the friendly Romanian group at DESY, especially *Dr. Nicoleta Baboi* and *Anca Ciobanu*. I thank them for their friendship, for our lovely "Romanian lunches" and their constant encouragement.

I would like to express my gratitude to my dear friends: *Ruxandra Lupu*, *Monica Dobre* and *Raluca Jucan*. I thank them for their support and for always being there for me.

My deepest, loving gratitude goes to Holger, to my parents, Elena and Ion and my grandparents, Maria and Ioan. None of this would have been possible without their love, patience and encouragement. I thank them for believing in me. I thank them for everything.

Appendix: Beam Energy Spectrum Effect on Cross-Sections

Data Samples	Sim.	$\sigma_{e_L^+e_R^-}$ [fb]	$\sigma_{e_R^+e_L^-}$ [fb]	$\sigma_{e_L^+e_L^-}$ [fb]	$\sigma_{e_R^+e_R^-}$ [fb]
$\tilde{\chi}_1^\pm$-hadronic	LoI	0.14	97.53	-	-
	DBD/SGV	0.18	93.69	-	-
$\tilde{\chi}_1^\pm$-semilept.	LoI	0.25	128.35	-	-
	DBD/SGV	0.25	120.6	-	-
$\tilde{\chi}_2^0$-hadronic	LoI	-	18.76	-	-
	DBD/SGV	-	16.69	-	-
$\tilde{\chi}_2^0$-semilept.	LoI	-	20.21	-	-
	DBD/SGV	-	19.47	-	-
SUSY bgrd.	LoI	50.16	103.51	-	-
	DBD/SGV	46.17	102.34	-	-
2f: Z-hadronic	LoI	18021.06	32317.27	-	-
	DBD/SGV	17994.7	32470.5	-	-
2f: Z-leptonic	LoI	4054.99	39626.55	-	-
	DBD/SGV	4344.86	5547.58	-	-
4f: WW-hadronic	LoI	29.94	6931.88	-	-
	DBD/SGV	33.52	7680.69	-	-
4f: WW-leptonic	LoI	12.36	3387.16	254.51	254.86
	DBD/SGV	12.54	3396.51	252.12	252
4f: WW-semilept.	LoI	67.23	17235	759.21	758.59
	DBD/SGV	68.41	17326.98	753.07	750.07
4f: ZZ-hadronic	LoI	273.22	660.89	-	-
	DBD/SGV	271.88	680.23	-	-
4f: ZZ-leptonic	LoI	85.93	285.18	64.07	63.93
	DBD/SGV	7159.37	7596.54	6991.01	6985
4f: ZZ-semilept.	LoI	2034.42	3469.04	1795.19	1795.9
	DBD/SGV	2073.89	3521.42	1775.49	1778.01
4f: WZ-hadronic	LoI	79.66	7047.38	-	-
	DBD/SGV	78.69	6400.11	-	-
4f: WZ-leptonic	LoI	42.43	1818.8	158.49	158.95
	DBD/SGV	47.57	1830.75	156.72	156.84
6f	LoI	596.1	1733.85	18.02	18.16
	DBD/SGV	716.77	1959.25	17.56	17.74

Table A.1: Table showing the difference in the cross-section values computed using the RDR (LoI) [54] and the TDR (DBD/SGV) [57] beam energy spectra. The shown numbers were computed with the Whizard event generator [167], considering 100% beam polarisation.

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation on my own and have not used other than the acknowledged resources and aids.

Madalina Chera
Hamburg, March 2018

Bibliography

[1] P.W. Higgs, "Broken symmetries, massless particles and gauge fields," *Phys. Lett.*, vol. 12, pp. 132–133, 1964. doi:10.1016/0031-9163(64)91136-9

[2] P.W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.*, vol. 13, p. 508, 1964. doi:10.1103/PhysRevLett.13.508

[3] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, p. 321, 1964. doi:10.1103/PhysRevLett.13.321

[4] The ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B*, vol. 716, Issue 1, pp. 1–29, 2012, arXiv:1207.7214v2 [hep-ex]. doi:10.1016/j.physletb.2012.08.020

[5] The CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett. B*, vol. 716, Issue 1, pp. 30–61, 2012, arXiv:1207.7235v2 [hep-ex]. doi:10.1016/j.physletb.2012.08.021

[6] Mikael Berggren, "SGV 3.0 - a fast detector simulation," 2008, arXiv:1203.0217v1 [physics.ins-det].

[7] M.A. Thomson , "Particle Flow Calorimetry and the PandoraPFA Algorithm," *Nucl. Instrum. Meth.*, vol. A611, pp. 25–46, 2009. doi:10.1016/j.nima.2009.09.009

[8] S. L. Glashow, "Partial Symmetries of Weak Interactions," *Nuclear Physics*, vol. 22, pp. 579–588, 1960. doi:10.1016/0029-5582(61)90469-2

[9] A. Salam, J.C. Ward, "Electromagnetic and Weak Interactions," *Physics Letters*, vol. 13, pp. 168–171, 1964. doi:10.1016/0031-9163(64)90711-5

[10] S. Weinberg, "A Model of Leptons," *Physical Review Letters*, vol. 19, p. 1264, 1967. doi:10.1103/PhysRevLett.19.1264

[11] C. Patrignani et al. (Particle Data Group), "2017 Review of Particle Physics," *Chin. Phys. C*, vol. 40, no. 100001, 2016 and 2017 update. http://pdg.lbl.gov/index.html

[12] B.P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), "Observation of Gravitational Waves from a Binary Black Hole Merger," *Phys. Rev. Lett.*, vol. 116, p. 061102, 2016, arXiv:1602.03837v1 [gr-qc]. doi:10.1103/PhysRevLett.116.061102

[13] The Particle Data Group, "Summary Table," http://pdg.lbl.gov/2017/tables/rpp2017-sum-gauge-higgs-bosons.pdf, 2017.

[14] H. Fritzsch, M. Gell-Mann and H.Leutwyler, "Advantages of the color octet gluon picture," *Physics Letter B*, vol. 47, pp. 365–368, 1973. doi:10.1016/0370-2693(73)90625-4

[15] E. Noether, "Invarianten beliebiger Differentialausdrücke," *Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen*, pp. 235–257, translation published in: Transport Theory and Statistical Physics 1(3) 1971, pp. 183-207.

[16] W.N. Cottingham and D.A. Greenwood, *An Introduction to the Standard Model of Particle Physics*, 2007, standard Model Lagrangian. doi:10.1017/CBO9780511791406

[17] J.A. Shifflett, "Standard Model Lagrangian," 2015, extracted from [16] and updated from summary tables published in 2015 by [11]. http://einstein-schrodinger.com/Standard_Model.pdf

[18] UA1 Collaboration, "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540GeV," *Phys. Lett. B*, vol. 122, pp. 103–116, 1983. doi:10.1016/0370-2693(83)91177-2

[19] UA1 Collaboration, "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider," *Phys. Lett. B*, vol. 126, pp. 398–410, 1983. doi:10.1016/0370-2693(83)90188-0

[20] UA2 Collaboration, "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN *pp* collider," *Phys. Lett. B*, vol. 122, pp. 476–485, 1983. doi:10.1016/0370-2693(83)91605-2

[21] UA2 Collaboration, "Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN *pp* collider," *Phys. Lett. B*, vol. 129, pp. 130–140, 1983. doi:10.1016/0370-2693(83)90744-X

[22] G.S. Guralnik, C.R. Hagen and T.W.B. Kibble, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, p. 585, 1964. doi:10.1103/PhysRevLett.13.585

[23] J. Goldstone, A. Salam, S. Weinberg, "Broken Symmetries," *Phys. Rev.*, vol. 127, p. 965, 1962. doi:10.1103/PhysRev.127.965

[24] G. Aad et al. (ATLAS Collaboration, CMS Collaboration), "Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s}$=7 and 8 TeV with the ATLAS and CMS Experiments," *Phys. Rev. Lett.*, vol. 114, p. 191803, 2015, arXiv:1503.07589v1 [hep-ex] . doi:10.1103/PhysRevLett.114.191803

[25] M. Gell-Mann, "A schematic model of baryons and mesons," *Phys. Lett.*, vol. 8, pp. 214–215, 1964. doi:10.1016/S0031-9163(64)92001-3

[26] G. Zweig, "An SU(3) Model for Strong Interaction Symmetry and its Breaking," *Developments in the Quark Theory of Hadrons*, vol. 1, pp. 22–101, 1964, CERN-TH-412. http://inspirehep.net/record/4674/files/cern-th-412.pdf

[27] R. Alkofer, J. Greensite, "Quark Confinement: The Hard Problem of Hadron Physics," *J. Phys. G*, vol. 34, 2007, arXiv:hep-ph/0610365v2. doi:10.1088/0954-3899/34/7/S02

[28] A. Ali, G. Kramer, "Jets and QCD: A Historical Review of the Discovery of the Quark and Gluon Jets and its Impact on QCD," *Eur. Phys. J. H*, vol. 36, p. 245, 2011, arXiv:1012.2288v2 [hep-ph]. doi:10.1140/epjh/e2011-10047-1

[29] B. Andersson, *The Lund Model*. Cambridge University Press, 1998.

[30] G. Marchesini, B.R. Webber, "Simulation of QCD Jets Including Soft Gluon Interference," *NUcl. Phys. B*, vol. 238, pp. 1–29, 1984, cERN-TH-3525. doi:10.1016/0550-3213(84)90463-2

[31] S.P. Martin, "A Supersymmetry Primer," *Perspectives on Supersymmetry*, vol. 18, 1998, arXiv:hep-ph/9709356v7. doi:10.1142/9789812839657_0001

[32] Andrei D. Sakharov, "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe," *Soviet Physics Uspekhi*, vol. 34, no. 5, p. 392, 1991. http://stacks.iop.org/0038-5670/34/i=5/a=A08

[33] M. Trodden, "Baryogenesis and Leptogenesis," 2004, arXiv:hep-ph/0411301v1.

[34] F. Capozzi et al., "Global constraints on absolute neutrino masses and their ordering," *Phys. Rev. D*, vol. 95, p. 096014, 2017, arXiv:1703.04471v1 [hep-ph]. doi:10.1103/PhysRevD.95.096014

[35] K.G. Begeman, A.H. Broeils, R.H. Sanders, "Extended rotation curves of spiral galaxies - Dark haloes and modified dynamics," *Monthly Notices of the Royal Astronomical Society*, vol. 249, pp. 523–537, 1991. doi:10.1093/mnras/249.3.523

[36] E. Komatsu, C.L. Bennett on behalf of the WMAP Science Team, "Results from the Wilkinson Microwave Anisotropy Probe," *Progress of Theoretical and Experimental Physics*, vol. 2014, no. 6, 2014, arXiv:1404.5415v2 [astro-ph.CO]. doi:10.1093/ptep/ptu083

[37] C. Patrignani et al. (Particle Data Group), "Supersymmetry, Part I (Theory)," *Chin. Phys. C*, vol. 40, 2017. http://pdg.lbl.gov/2017/reviews/rpp2017-rev-susy-1-theory.pdf

[38] H. Baer, X. Tata, "Weak Scale Supersymmetry : From Superfields to Scattering Events," 2012, Cambridge University Press.

[39] S. Choi, J. Kalinowski, G. MoortgatâĂŞPick et al., "Analysis of the neutralino system in supersymmetric theories," *Eur. Phys. J. C*, vol. 22, pp. 563–579, 2001, arXiv:hep-ph/0108117v1. doi:10.1007/s100520100808

[40] J. Kalinowski, G. MoortgatâĂŞPick, "Complete Reconstruction of the Neutralino System," *eConf C010630 (2001) P323*, arXiv:hep-ph/0202083v1.

[41] A. H. Chamseddine, R. Arnowitt, P. Nath, "Locally Supersymmetric Grand Unification," *Phys. Rev. Lett.*, vol. 49, p. 970, 1982. doi:10.1103/PhysRevLett.49.970

[42] R.Barbieri, S.Ferrara, C.A.Savoy, "Gauge models with spontaneously broken local supersymmetry," *Phys. Lett. B*, vol. 119, pp. 343–347, 1982. doi:10.1016/0370-2693(82)90685-2

[43] M. Dine, W. Fischler, "A Phenomenological Model of Particle Physics Based on Supersymmetry ," *Phys. Lett. B*, vol. 110, pp. 227–331, 1981, pRINT-82-0029 (IAS,PRINCETON). doi:10.1016/0370-2693(82)91241-2

[44] C.R. Nappi, B.A. Ovrut, "Supersymmetric Extension of the SU(3) x SU(2) x U(1) Model," *Phys. Lett. B*, vol. 113, pp. 175–179, 1982, pRINT-82-0093 (IAS,PRINCETON). doi:10.1016/0370-2693(82)90418-X

[45] L. Randall, R. Sundrum, "Out of this world supersymmetry breaking," *Nucl. Phys. B*, vol. 557, pp. 79–118, 1999. doi:10.1016/S0550-3213(99)00359-4

[46] G.F. Giudice, R. Rattazzi, M.A. Luty, H. Murayama, "Gaugino mass without singlets," *JHEP*, vol. 1998, no. 12, p. 027, 1999. http://stacks.iop.org/1126-6708/1998/i=12/a=027

[47] The ATLAS Collaboration, G. Aad, E. Abat, J Abdallah et al., "The ATLAS Experiment at the CERN Large Hadron Collider," *Journal of Instrumentation*, vol. 3, 2008. doi:10.1088/1748-0221/3/08/S08003

[48] The CMS Collaboration, S Chatrchyan et al., "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, p. S08004, 2008. http://iopscience.iop.org/1748-0221/3/08/S08004

[49] T. Behnke, J. E. Brau, B. Foster et al., "The International Linear Collider Technical Design Report - Volume I: Executive Summary," 2013, DESY 13-062, CERN-ATS-2013-037.

[50] K. Fuji, C. Grojean, M.E. Peskin et al., "Physics Case for the International Linear Collider," DESY 15-094, 2015, arXiv:1506.05992v2 [hep-ex].

[51] O. Buchmueller, P. de Jong, "Supersymmetry, Part II (Experiment)," 2017, PDG Summary review. http://pdg.lbl.gov/2017/reviews/rpp2017-rev-susy-2-experiment.pdf

[52] M. Aicheler, P. Burrows, M. Draper, T. Garvey, P. Lebrun, K. Peach, N. Phinney, H. Schmickler, D. Schulte and N. Toge, "A Multi-TeV linear collider based on CLIC technology: CLIC Conceptual Design Report," 2012, CERN-2012-007.

[53] "The Linear Collider Collaboration." https://www.linearcollider.org/

[54] J. Brau, Y. Okada, N. Walker et al., "ILC Reference Design Report Volume 1 - Executive Summary," 2007, arXiv:0712.1950v1 [physics.acc-ph].

[55] Nan Phinney, Nobukasu Toge, Nicholas Walker et al., "ILC Reference Design Report Volume 3 - Accelerator," 2007, arXiv:0712.2361 [physics.acc-ph].

[56] Chris Adolphsen, Maura Barone, Barry Barish, Karsten Buesser et al., "The International Linear Collider Technical Design Report - Volume 3.I: Accelerator R&D in the Technical Design Phase," 2013, CERN-ATS-2013-037, DESY 13-062.

[57] Chris Adolphsen, Maura Barone, Barry Barish, Karsten Buesser et al., "The International Linear Collider Technical Design Report - Volume 3.II: Accelerator Baseline Design," 2013, arXiv:1306.6328 [physics.acc-ph].

[58] The TASSO Collaboration, "Evidence for a spin-1 gluon in three-jet events," *Physics Letters B*, vol. 97, Issues 3-4, pp. 453–458, 1980. doi:10.1016/0370-2693(80)90639-5

[59] The PLUTO Collaboration, "Evidence for gluon bremsstrahlung in e^+e^- annihilations at high energies," *Physics Letters B*, vol. 86, Issues 3-4, pp. 418–425, 1979. doi:10.1016/0370-2693(79)90869-4

[60] C. Campagnari, M. Franklin, "The discovery of the top quark," *Reviews of Modern Physics*, vol. 69, Issue 1, pp. 137–211, 1997. doi:10.1103/RevModPhys.69.137

[61] The CDF Collaboration, "Observation of Top Quark Production in $p\overline{p}$ Collisions with the Collider Detector at Fermilab," *Phys. Rev. Lett.*, vol. 74, Issue 14, p. 2626, 1995.

[62] The D0 Collaboration, "Observation of the Top Quark," *Phys. Rev. Lett.*, vol. 74, Issue 14, p. 2632, 1995.

[63] M. Tigner, "A Possible Apparatus for Electron Clashing-Beam Experiments," *Nuovo Cim.*, vol. 37, Issue 3, pp. 1228–1231, 1965. doi:10.1007/BF02773204

[64] B. Richter, "Very high energy electron-positron colliding beams for the study of weak interactions," *Nuclear Instruments and Methods*, vol. 136, Issue 1, pp. 47–60, 1976. doi:10.1016/0029-554X(76)90396-7

[65] "The European XFEL." http://www.xfel.eu/en/

[66] A. Vogel, "Beam-induced backgrounds in detectors at the ILC," Ph.D. dissertation, University of Hamburg, 2008, DESY-THESIS-2008-036.

[67] K. Yokoya, P. Chen, "Beam-beam phenomena in linear colliders," *Lect. Notes Phys.*, vol. 400, pp. 415–445, 1992. doi:10.1007/3-540-55250-2_37

[68] C. Rimbault, P. Bambade, K. Mönig, and D. Schulte, "Incoherent pair generation in a beam-beam interaction simulation," *Phys. Rev. ST Accel. Beams*, vol. 9, p. 034402, 2006. doi:10.1103/PhysRevSTAB.9.034402. https://link.aps.org/doi/10.1103/PhysRevSTAB.9.034402

[69] T. Sjostrand, S. Mrenna, P. Skands, "PYTHIA 6.4 Physics and Manual," 2006.

[70] Howard Baer, Tim Barklow, Keisuke Fuji, Yuanning Gao et al., "The International Linear Collider Technical Design Report - Volume II: Physics," 2013, arXiv:1306.6352v1 [hep-ph].

[71] G. Moortgat-Pick, H. Baer, M. Battaglia et al., "Physics at the e^+e^- linear collider," *Eur. Phys. Journal C*, vol. 75, p. 371, 2015, 1504.01726 [hep-ph]. doi:10.1140/epjc/s10052-015-3511-9

[72] C.F. Dürig, "Measuring the Higgs Self-coupling at the International Linear Collider," dESY-THESIS-2016-027. doi:10.3204/PUBDB-2016-04283

[73] J. Tian, "Study of Higgs self-coupling at the ILC based on the full detector simulation at $\sqrt{s}$=500 GeV and $\sqrt{s}$=1 TeV," 2013, LC-REP-2013-003. http://www-flc.desy.de/lcnotes/notes/LC-REP-2013-003.pdf

[74] G. Degrassi, S. Di Vita et al., "Higgs mass and vacuum stability in the Standard Model at NNLO," *J. High Energ. Phys.*, 2012, arXiv:1205.6497v2 [hep-ph]. doi:10.1007/JHEP08(2012)098

[75] M. Beneke, M. Steinhauser, "Non-relativistic high-energy physics: top production and dark matter annihilation," *Nucl. Part. Phys. Proc.*, vol. 261-262, pp. 378–413, 2015, arXiv:1506.07962v1 [hep-ph]. doi:10.1016/j.nuclphysbps.2015.03.024

[76] K. Seidel, F. Simon et al., "Top quark mass measurements at and above threshold at CLIC," *Eur. Phys. J. C*, vol. 73, p. 2530, 2013. doi:10.1140/epjc/s10052-013-2530-7

[77] The CMS Collaboration, "ECFA 2016: Prospects for selected standard model measurements with the CMS experiment at the High-Luminosity LHC," 2016, CMS-PAS-FTR-16-006. http://cds.cern.ch/record/2262606/files/FTR-16-006-pas.pdf

[78] LCC Physics Working Group - K. Fuji et al., "The Potential of the ILC for Discovering New Particles," 2016, DESY 17-012, arXiv:1702.05333v1 [hep-ph].

[79] M. Habermehl, K. Fuji, J. List et al., "WIMP Searches at the International Linear Collider," 2017, arXiv:1702.05377v1 [hep-ex].

[80] Berggren, M., Brümmer, F., List, J. et al., "Tackling light higgsinos at the ILC," *Eur. Phys. J.*, vol. C 73:2660, 2013, arXiv:1307.3566v2 [hep-ph]. doi:10.1140/epjc/s10052-013-2660-y

[81] ILC Parameters Joint Working Group - T. Barklow et al., "ILC Operating Scenarios," DESY 15-102, 2015, arXiv:1506.07830v1 [hep-ex].

[82] L. Evans, S. Michizono, "The International Linear Collider Machine Staging Report 2017," 2017, arXiv:1711.00568v3 [physics.acc-ph].

[83] G. Hanson et al., "Evidence for Jet Structure in Hadron Production by e^+e^- Annihilation," *Phys. Rev. Lett.*, vol. 35, pp. 1609–1612, 1975. doi:10.1103/PhysRevLett.35.1609. https://link.aps.org/doi/10.1103/PhysRevLett.35.1609

[84] G. Hanson, "Hadron production by e^+e^- annihilation at center-of-mass energies between 2.6 and 7.8 GeV. II. Jet structure and related inclusive distributions," *Phys. Rev. D*, vol. 26, pp. 991–1012, 1982. doi:10.1103/PhysRevD.26.991. https://link.aps.org/doi/10.1103/PhysRevD.26.991

[85] Ch. Berger et al., "A study of jets in electron positron annihilation into hadrons in the energy range 3.1 to 9.5 GeV," *Physics Letters B*, vol. 78, no. 1, pp. 176–182, 1978. doi:https://doi.org/10.1016/0370-2693(78)90377-5. http://www.sciencedirect.com/science/article/pii/0370269378903775

[86] J. E. Huth et al., "Towards a standardization of jet definitions," in *Research Directions for the Decade*. http://lss.fnal.gov/archive/1990/conf/Conf-90-249-E.pdf

[87] G.P. Salam, "Towards Jetography," *The European Physical Journal C*, vol. 67, pp. 637–686, arXiv:0906.1833v2 [hep-ph]. doi:https://doi.org/10.1140/epjc/s10052-010-1314-6

[88] G.P. Salam, G. Soyez, "A Practical Seedless Infrared-Safe Cone Jet Algorithm, journal = Journal of High Energy Physics ," vol. 2007, no. 05, p. 086, 2007. http://stacks.iop.org/1126-6708/2007/i=05/a=086

[89] S. Catani, U.L. Dokshitzer, M. Olsson, G. Turnock, B.R. Webber, "New clustering algorithm for multijet cross sections in e^+e^- annihilation," *Physics Letters B*, vol. 269, no. 3, pp. 432–438, 1991. doi:https://doi.org/10.1016/0370-2693(91)90196-W. http://www.sciencedirect.com/science/article/pii/037026939190196W

[90] Yu.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, "Better jet clustering algorithms," *Journal of High Energy Physics*, vol. 1997, no. 08, p. 001, arXiv:hep-ph/9707323v2. http://stacks.iop.org/1126-6708/1997/i=08/a=001

[91] M. Berggren, "Truth Algorithms - DBD Status and Updates," DESY, Presentation at the "High Level Reconstruction Workshop". https://agenda.linearcollider.org/event/6787/contributions/33406/attachments/27501/41765/berggren-HLRecoWS-desy-jul2015.pdf

[92] D. Buskulic et al., "Performance of the ALEPH Detector at LEP," *Nucl. Instrum. Meth. in Phys. Res.*, vol. A360, pp. 481–506, 1995. doi:doi:10.1016/0168-9002(95)00138-7

[93] M. Derrick et al., "Design and construction of the ZEUS barrel calorimeter," *Nucl. Instrum. Meth. in Phys. Res. A*, vol. 309, pp. 77–100, 1991.

[94] The ZEUS Calorimeter Group, J.A. Crittenden, "The Performance of the ZEUS Calorimeter," 1994.

[95] H1 Collaboration: I. Abt et al., "The H1 detector at HERA," *Nucl. Instrum. Meth. A*, vol. 386, pp. 310–347, 1997. doi:10.1016/S0168-9002(96)00893-5

[96] H1 Collaboration: I. Abt et al., "The tracking, calorimeter and muon detectors of the H1 experiment at HERA," *Nucl. Instrum. Meth. A*, vol. 386, pp. 348–396, 1997. doi:10.1016/S0168-9002(96)00894-7

[97] ATLAS Collaboration: G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider," *Journal of Instrumentation*, vol. 3, p. S08003, 2008. http://stacks.iop.org/1748-0221/3/i=08/a=S08003

[98] The CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s}$=7 TeV," *Journal of Instrumentation*, vol. 8, p. P09009, 2013, arXiv:1306.2016v2 [hep-ex]. doi:10.1088/1748-0221/8/09/P09009

[99] C. Amsler et al. (Particle Data Group), "Review of Particle Physics," *Physics Letters B* , vol. 667, 2008.

[100] A. Ebrahimi, "Jet Energy Measurements at ILC: Calorimeter DAQ Requirements and Application in Higgs Boson Mass Measurements," PhD Dissertation, DESY-THESIS-2017-045. https://bib-pubdb1.desy.de/record/394104/

[101] Junping Tian, Keisuke Fujii, "Measurement of Higgs couplings and self-coupling at the ILC," 2013, arXiv:1311.6528v1 [hep-ph].

[102] S. Catani, Yu.L. Dokshitzer, M. Olsson et al., "New clustering algorithm for multijet cross sections in e^+e^- annihilation," *Phys. Lett. B*, vol. 269, pp. 432–438, 1991. doi:10.1016/0370-2693(91)90196-W

[103] D.Decamp et al., ALEPH Collaboration, "ALEPH: A detector for electron-positron annihilations at LEP," *Nucl. Instrum. Meth. in Phys. Res.*, vol. A294, pp. 121–178, 1990. doi:10.1016/0168-9002(90)91831-U

[104] The ZEUS Collaboration, "The ZEUS Detector," Status Report, 1993. http://www-zeus.desy.de/bluebook/scanned_bluebook.pdf

[105] F. Abe et al., "The CDF Detector: An Overview," *Nucl. Instrum. Meth. in Phys. Res.*, vol. A271, pp. 387–403, 1988. doi:10.1016/0168-9002(88)90298-7

[106] Patrick Janot, "Particle-Flow Event Reconstruction from LEP to LHC," Lecture at the EDIT School, CERN, 31 Jan - 10 Feb 2011. https://indico.cern.ch/event/96989/contributions/2124494/attachments/1114188/1589704/ParticleFlow.pdf

[107] I.G. Knowles5 and G. D. Lafferty, "Hadronization in Z^0 decay," *Journal of Physics G: Nuclear and Particle Physics*, vol. G23, pp. 731–789, 1997. doi:10.1088/0954-3899/23/7/003

[108] B. Hermberg, "Measuring Hadronic Jets at the ILC - From Particle Flow Calorimetry to the Higgs Self-Coupling," Ph.D. dissertation, University of Hamburg, 2015, DESY-THESIS-2015-039.

[109] The ILD Concept Group, "The International Large Detector Letter of Intent," 2009, arXiv:1006.3396.

[110] http://aleph.web.cern.ch/aleph/aleph/alephgif/aleph1.gif

[111] CDF Run II Collaboration.
https://www-cdf.fnal.gov/upgrades/upgrades.html

[112] David Barney.
https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMS

[113] W.R. Leo, *Techniques for Nuclear and Particle Physics Experiments.* Springer Verlag, 1993.

[114] *CDF Calorimeter and Its Upgrade*, FERMILAB-Conf-95/001-E.

[115] L. Balka et al., "The CDF Electromagnetic Calorimeter," *Nucl. Instrum. Meth. in Phys. Res. A*, vol. A267, pp. 272–279, 1988. doi:10.1016/0168-9002(88)90474-3

[116] G.M. Briskin, "Diffractive Dissociation in ep Deep Inelastic Scattering," Ph.D. dissertation, Tel Aviv University, School of Physics and Astronomy, 1998, DESY-THESIS-1998-036.

[117] Andrea Bocci, "A Study of Jet Energy Measuremet at CDF," Master's thesis, University of Pisa, The Faculty of Mathematical, Physical and Natural Science, 1998.
http://lss.fnal.gov/archive/thesis/1900/fermilab-thesis-1998-04.pdf

[118] A. Bocci, S. Kuhlmann, S. Lami, G. Latino, "Study of Jet Energy Resolution At CDF," *International Journal of Modern Physics A*, vol. 16, pp. 255–258, 2001. doi:10.1142/S0217751X01006632

[119] The CMS Collaboration, "Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}," 2009, CMS PAS PFT-09/001.
https://cds.cern.ch/record/1194487

[120] F. Abe et al., "Study of four-jet events and evidence for double parton interactions in $p\bar{p}$ collisions at $\sqrt{s}$=1.8 TeV," *Phys. Rev. D*, vol. 47, pp. 4857–4871, 1993. doi:10.1103/PhysRevD.47.4857

[121] The CMS Collaboration, "Jet Energy Resolution in CMS at $\sqrt{s}$=7 TeV," 2011. https://cds.cern.ch/record/1339945/files/JME-10-014-pas.pdf

[122] F. Beaudette, on behalf of the CMS Collaboration, "The CMS Particle Flow Algorithm," 2014, arXiv:1401.8155v1 [hep-ex].

[123] Ties Behnke, James E. Brau, Philip Burrows, Juan Fuster et al., "The International Linear Collider Technical Design Report - Volume IV: Detectors," 2013, arXiv:1306.6329v1 [physics.ins-det].

[124] I. Valin, C. Hu-Guo, J. Baudot et al., "A reticle size CMOS pixel sensor dedicated to the STAR HFT," *Journal of Instrumentation*, vol. 7, no. 01, p. C01102, 2012. http://stacks.iop.org/1748-0221/7/i=01/a=C01102

[125] O. Alonso, R. Casanova, A. Dieguez et al., "DEPFET active pixel detectors for a future linear e^+e^- collider," 2012, arXiv:1212.2160v1 [physics.ins-det].

[126] L. Greiner, E. Anderssen, H.S. Matis et al., "A MAPS based vertex detector for the STAR experiment at RHIC," *Nucl. Instrum. Meth.*, vol. A650, no. 1, pp. 68 – 72, 2011. http://www.sciencedirect.com/science/article/pii/S0168900210027439

[127] Aurore Savoy-Navarro, "Large Area Silicon Tracking: New Perspectives," 2012, arXiv:1203.0736v1 [hep-ex].

[128] Aurore Savoy-Navarro, "Overview of the SiLC R&D Activities," 2009, arXiv:0903.4361v1 [physics.ins-det].

[129] W.B. Atwood et al., "Performance of the ALEPH Time Projection Chamber," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 306(3), pp. 446–458, 1991. doi:10.1016/0168-9002(91)90038-R

[130] The LCTPC Collaboration, http://www.lctpc.org.

[131] J. Beringer et al. (Particle Data Group), "The Review of Particle Physics," *Phys. Rev. D86*, 010001 (2012).

[132] Y. Giomataris, P. Rebourgeard, J. P. Robert, and G. Charpak, "MICROMEGAS: a high-granularity position-sensitive gaseous detector for high particle-flux environments," *Nucl. Instrum. Meth.*, vol. A376, pp. 29–35, 1996.

[133] F. Sauli, "GEM: A new concept for electron amplification in gas detectors," *Nucl. Instrum. Meth.*, vol. A386, pp. 531–534, 1997.

[134] LCTPC Collaboration, "Report to the DESY PRC 2010," *ILC NOTE 2012-063, LC-DET-2012-067*, pp. 3–5.

[135] T. Lux, "Studies for a Time Projection Chamber for the Internationnal Linear Collider and Measurement of Beauty Cross Section in Deep Inelastic Scattering at HERA," Ph.D. dissertation, University of Hamburg, 2005, DESY-THESIS-2005-019.

[136] M. Kobayashi et al., "Cosmic ray tests of a GEM-based TPC prototype operated in Ar-CF4-isobutane gas mixtures," *Nucl. Instrum. Meth.*, vol. A641, pp. 37–47, 2011, arXiv:1008.5068v2 [physics.ins-det].

[137] The CALICE Collaboration. https://twiki.cern.ch/twiki/bin/view/CALICE/WebHome

[138] The CALICE collaboration, C.Adloff et al., "Response of the CALICE Si-W Electromagnetic Calorimeter Physics Prototype to Electrons," *Nuclear Instruments and Methods in Physics Research*, vol. 608, no. 3, pp. 372–383, 2009, arXiv:0811.2354v1 [physics.ins-det]. doi:10.1016/j.nima.2009.07.026

[139] http://www.linearcollider.org/ILC/Press/Images-and-graphics

[140] The CALICE collaboration, C. Adloff et al., "Hadronic energy resolution of a highly granular scintillator-steel hadron calorimeter using software compensation techniques," *Journal of Instrumentation*, vol. 7, p. P09017, 2012, arXiv:1207.4210v2 [physics.ins-det]. doi:10.1088/1748-0221/7/09/P09017

[141] M. Czakon, J. Gluza, T. Riemann, "The planar four-point master integrals for massive two-loop Bhabha scattering," *Nuclear Physics B*, vol. 751, pp. 1–17, 2006, arXiv:hep-ph/0604101v1. doi:10.1016/j.nuclphysb.2006.05.033

[142] H. Abramowicz, A. Abusleme, K. Afanaciev et al., "Forward Instrumentation for ILC Detectors," *Journal of Instrumentation*, vol. 5, p. P12002, 2010, arXiv:1009.2433v2 [physics.ins-det]. doi:10.1088/1748-0221/5/12/P12002

[143] F. Kircher et al., "CMS Coil Design and Assembly," *IEEE Transactions on Applied Superconductivity*, vol. 12, pp. 395–398, 2002. doi:10.1109/TASC.2002.1018427

[144] F. Kircher, U. Schneekloth et al., "Conceptual Design of the ILD Detector Magnet System," 2002, lC-DET-2012-081.

[145] The CALICE Collaboration, C. Adloff et al., "Construction and performance of a silicon photomultiplier/extruded scintillator tail-catcher and muon-tracker," *Journal of Instrumentation*, vol. 7, p. P04015, 2012, arXiv:1201.1653v4 [physics.ins-det]. doi:10.1088/1748-0221/7/04/P04015

[146] "ILC Soft." http://ilcsoft.desy.de/portal

[147] F. Gaede, T. Behnke, N. Graf, T. Johnson, "LCIO - A persistency framework for linear collider simulation studies," 2003, arXiv:physics/0306114v1.

[148] S. Aplin, J. Engels, F. Gaede, N. A. Graf, T. Johnson, J. McCormick, "LCIO: A Persistency Framework and Event Data Model for HEP," 2012, SLAC-PUB-15296.

[149] "The Whizard Event Generator." https://whizard.hepforge.org/

[150] B. C. Allanach, C. Balazs, G. Belanger et al., "SUSY Les Houches Accord 2," *Comp.Phys.Commun.*, vol. 180, pp. 8–25, 2009, arXiv:0801.0045v3 [hep-ph]. doi:10.1016/j.cpc.2008.08.004

[151] D. Schulte, "Study of electromagnetic and hadronic background in the interaction region of the TESLA Collider," dESY-TESLA-97-08.

[152] M. Moretti, T. Ohl and J. Reuter , "O'Mega: An Optimizing Matrix Element Generator," 2001, LC-TOOL-2001-040.

[153] M. Berggren, "Pythia Tune," presentation given in the "General Meetings of CLICdp WG Analysis". https://indico.cern.ch/event/113752/contributions/1321047/attachments/44245/63700/pythiatune.pdf

[154] T. Sjöstrand, S. Mrenna and P. Skands, "PYTHIA 6.4 physics and manual," *Journal of High Energy Physics*, 2006, arXiv:hep-ph/0603175v2. doi:10.1088/1126-6708/2006/05/026

[155] GEANT4 Collaboration, "Geant 4 - A Simulation Toolkit," 2003.

[156] P. Mora de Freitas, H. Videau, "Detector Simulation with Mokka/Geant4: Present and Future," 2003, Conference Note for the International Workshop on Linear Colliders (LCWS 2002).

[157] "MySQL Homepage." https://www.mysql.com/

[158] "ILD Reference Detector Model." http://www.ilcild.org/documents/ild-letter-of-intent/ild-loi-material/ILDreferencedetector.pdf/view

[159] "MARLIN - Modular Analysis and Reconstruction for the Linear Collider." http://ilcsoft.desy.de/portal/software_packages/marlin/index_eng.html

[160] "GEAR - A Geometry Description Toolkit." http://ilcsoft.desy.de/portal/software_packages/gear

[161] LCFI Collaboration, "The LCFIVertex Package: Vertexing, Flavour Tagging and Vertex Charge Reconstruction with an ILC Vertex Detector," *Nucl. Instrum. Meth.*, vol. A610, pp. 573–589, 2009, arXiv:0908.3019v1. doi:10.1016/j.nima.2009.08.059

[162] D. J. Jackson, "A Topological Vertex Reconstruction Algorithm for Hadronic Jets," *Nucl. Instrum. Meth.*, vol. A388, pp. 247–253, 1997. doi:10.1016/S0168-9002(97)00341-0

[163] J.S. Marshall, A. Münnich, M.A. Thomson , "Performance of Particle Flow Calorimetry at CLIC," *Nucl. Instrum. Meth.*, vol. A700, pp. 153–162, 2013, arXiv:1209.4039v1 [physics.ins-det]. doi:10.1016/j.nima.2012.10.038

[164] J.S. Marshall, M.A. Thomson, "The Pandora Software Development Kit for Pattern Recognition," *Eur. Phys. J. C*, vol. 75, p. 439, 2015, arXiv:1506.05348v2 [physics.data-an]. doi:10.1140/epjc/s10052-015-3659-3

[165] B. Xu, "Improvement of photon reconstruction in PandoraPFA," 2016, Talk presented at the International Workshop on Future Linear Colliders (LCWS15), Whistler, Canada, 2-6 November 2015, arXiv:1603.00013v1 [physics.ins-det].

[166] B. Xu, "Detectors and Physics at a Future Linear Collider," 2017. doi:10.17863/CAM.16867

[167] W. Kilian, T. Ohl and J. Reuter , "WHIZARD simulating multi-particle processes at LHC and ILC," *Eur. Phys. J.*, vol. C 71, 2011, arXiv:0708.4233v2 [hep-ph]. doi:10.1140/epjc/s10052-011-1742-y

[168] Andrea Giammanco, "The Fast Simulation of the CMS Experiment," *Journal of Physics: Conference Series*, vol. 513, p. 022012, 2014, Contribution to ICHEP2013. doi:10.1088/1742-6596/513/2/022012

[169] W. Lukas, The ATLAS Collaboration, "Fast Simulation for ATLAS: Atlfast-II and ISF," *Journal of Physics: Conference Series*, vol. 396, p. 022031, 2012. doi:10.1088/1742-6596/396/2/022031

[170] J. de Favereau, C. Delaere, P. Demin et al., "DELPHES 3 - A modular framework for fast simulation of a generic collider experiment," 2014, arXiv:1307.6346v3 [hep-ex].

[171] The DELPHI Collaboration, P. Abreu et al., "Performance of the DELPHI detector," *Nuclear Instntments and Methods in Physics Research* , vol. A378, pp. 57–100, 1996. doi:10.1016/0168-9002(96)00463-9

[172] S. Schreiber, "THE TESLA LINEAR COLLIDER PROJECT," *International Journal of Modern Physics A*, vol. 13, no. 14, pp. 2431–2441, 1998. doi:10.1142/S0217751X98001220

[173] S. Jain, The CMS Collaboration, "CMS Fast Simulation: A Tool for Physics Searches at LHC," *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 317–321, 2010. doi:10.1109/NSSMIC.2010.5873772

[174] R. Rahmat, R. Kroeger, A. Giammanco, "The Fast Simulation of the CMS Experiment," *Journal of Physics: Conference Series*, vol. 396, p. 062016, 2012, Contribution to ICHEP2012. doi:10.1088/1742-6596/396/6/062016

[175] The ATLAS Collaboration, "The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim," 2010, aTL-PHYS-PUB-2010-013, ATL-COM-PHYS-2010-838.

[176] E. Barberio, J. Boudreau, B. Butler et al., "Fast Simulation of Electromagnetic Showers in the ATLAS Calorimeter: Frozen Showers," *Journal of Physics: Conference Series*, vol. 160, p. 012082, 2009. doi:10.1088/1742-6596/160/1/012082

[177] M. Selvaggi, The DELPHES Team, "DELPHES Fast Simulation," 2014, Presentation at the Second Fast Monte Carlo in HEP Workshop. https://indico.cern.ch/event/280658/contributions/638034/attachments/514718/710220/delphes_mem.pdf

[178] M. Berggren, "SGV Reference Manual." http://delphiwww.cern.ch/~berggren/sgv_refman/sgv_refman.html

[179] K. Edmonds, S. Fleischmann, T. Lenz et al., "The fast ATLAS track simulation (FATRAS)," 2008, aTL-SOFT-PUB-2008-001, ATL-COM-SOFT-2008-002.

[180] The CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *Journal of Instrumentation*, vol. 9, 2014. http://iopscience.iop.org/1748-0221/9/10/P10009

[181] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960. doi:10.1115/1.3662552

[182] G. Grindhammer, M. Rudowicz, S. Peters, "The fast simulation of electromagnetic and hadronic showers," *Nucl. Instrum. Menth*, vol. A290, pp. 469–488, 1990, SLAC-PUB-5072. doi:10.1016/0168-9002(90)90566-O

[183] G. Grindhammer, S. Peters, "The Parameterized Simulation of Electromagnetic Showers in Homogeneous and Sampling Calorimeters," contribution at Int. Conf. on Monte Carlo Simulation in High Energy and Nuclear Physics, 1993.

[184] Matteo Cacciari, Gavin P. Salam, Gregory Soyez, "FastJet User Manual," 2011, arXiv:1111.6097v1 [hep-ph].

[185] M. Berggren, "SGV - a fast detector simulation," Talk at the CLIC Workshop, CERN, January, 2013. https://indico.cern.ch/event/204269/contributions/1492132/attachments/309354/431911/berggren-sgv-clicws-jan2013.pdf

[186] T. Krämer, "Track Parameters in LCIO," 2006, LC-DET-2006-004.

[187] P. Billoir, "Track Fitting With Multiple Scattering: A New Method," *Nuclear Instruments and Methods in Physics*, vol. 225, pp. 352–366, 1984. doi:10.1016/0167-5087(84)90274-6

[188] P. Billoir, "Progressive track recognition with a Kalman-like fitting procedure," *Computer Physics Communications*, vol. 57, pp. 390–394, 1989. doi:10.1016/0010-4655(89)90249-X

[189] R. Früwirth, "Application of Kalman Filtering to Track and Vertex Finding," *Nucl. Instrum. and Meth.*, vol. A262, pp. 444–450, 1987. doi:10.1016/0168-9002(87)90887-4

[190] G.H Golub, C.F. Van Loan, *Matrix Computations*, 3rd ed. John Hopkins University Press.

[191] A. Vogel, "The Coordinate System for LDC Detector Studies," 2005, LC-DET-2005-009. http://flc.desy.de/lcnotes/notes/localfsExplorer_read?currentPath=/afs/desy.de/group/flc/lcnotes/LC-DET-2005-009.pdf

[192] T. Krämer, "Track Parameters in LCIO," 2006, lC-DET-2006-004. http://flc.desy.de/lcnotes/notes/localfsExplorer_read?currentPath=/afs/desy.de/group/flc/lcnotes/LC-DET-2006-004.pdf

[193] J. Alcaraz, "Helicoidal tracks," 1995, l3 Internal Note 1666. http://www.lcsim.org/sites/lcsim/apidocs/org/lcsim/util/swim/doc-files/L3_helix.pdf

[194] P. Bourke, "Equation of a circle from 3 points," 1990. http://paulbourke.net/geometry/circlesphere/

[195] M. Battaglia, T. Barklow, M. Peskin, Y. Okada, S. Yamashita, P. Zerwas, "Physics Benchmarks for the ILC Detectors," 2006, report of the ILC Benchmark panel, arXiv:hep-ex/0603010v1.

[196] T. Suehara, J. List, "Chargino and Neutralino Separation with the ILD Experiment," 2009, arXiv:0906.5508 [hep-ex].

[197] Y. Li, A. Nomerotski, "Chargino and Neutralino Masses at ILC," 2010, proceedings for the Linear Collider Workshop 2010, arXiv:1007.0698v1 [physics.ins-det].

[198] H. Aihara, P. Burrows, M. Oreglia et al., "SiD Letter of Intent," 2009, Letter of Intent for SiD detector concept presented to ILC IDAG, arXiv:0911.0006v1 [physics.ins-det].

[199] A. De Roeck, J. Ellis, F. Gianotti, F. Moortgat, K.A. Olive, L. Pape, "Supersymmetric benchmarks with non-universal scalar masses or gravitino dark matter," *Eur. Phys. J C*, vol. 49, p. 1041, 2007. doi:10.1140/epjc/s10052-006-0182-6

[200] W. Porod, "SPheno, a program for calculating supersymmetric spectra, SUSY particle decays and SUSY particle production at e^+e^- colliders," *Comput. Phys. Commun.*, vol. 153, 2003, arXiv:hep-ph/0301101v4 . doi:10.1016/S0010-4655(03)00222-4

[201] W. Porod, F. Staub, "`SPheno 3.1`: extensions including flavour, CP-phases and models beyond the MSSM," *Comput. Phys. Commun.*, vol. 183, 2012, arXiv:1104.1573v3 [hep-ph]. doi:10.1016/j.cpc.2012.05.021

[202] Sneutrino IdentiiňĄcation in Lepton Pair Production at ILC with Polarized Beams, "A. V. Tsytrinov, J. Kalinowski, P. Osland, A. A. Pankov," *Phys. Lett. B*, vol. 718, pp. 94–99, 2012, arXiv:1207.6234v1 [hep-ph]. doi:10.1016/j.physletb.2012.10.007

[203] The ATLAS Collaboration. https://atlas.web.cern.ch/Atlas/Collaboration/

[204] The CMS Collaboration. http://cms.web.cern.ch/content/cms-collaboration

[205] J. Alwall, P. Schuster, N. Toro, "Simplified Models for a First Characterization of New Physics at the LHC," *Phys.Rev.D*, vol. 79, p. 075020, 2009, arXiv:0810.3921v2 [hep-ph]. doi:10.1103/PhysRevD.79.075020

[206] D. Alves et al., "Simplified Models for LHC New Physics Searches," *J. Phys. G: Nucl. Part. Phys.*, vol. 39, p. 105005, 2012, arXiv:1105.2838v1 [hep-ph]. doi:10.1088/0954-3899/39/10/105005

[207] CMS Collaboration, "Interpretation of searches for supersymmetry with simplified models," *Phys. Rev. D*, vol. 88, p. 052017, 2013, arXiv:1301.2175v2 [hep-ex]. doi:10.1103/PhysRevD.88.052017

[208] The ATLAS Colaboration, "Search for electroweak production of supersymmetric particles in the two and three lepton final state at $\sqrt{13}$ TeV with the ATLAS detector," 2017, ATLAS-CONF-2017-039. http://cds.cern.ch/record/2267406/files/ATLAS-CONF-2017-039.pdf

[209] The CMS Collaboration, "Combined search for electroweak production of charginos and neutralinos in pp collisions at $\sqrt{13}$ TeV," 2017, CMS-PAS-SUS-17-004. http://cds.cern.ch/record/2273907/files/SUS-17-004-pas.pdf

[210] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *J. High Energ. Phys.*, vol. 2014, p. 79, 2014, arXiv:1405.0301v2 [hep-ph]. doi:10.1007/JHEP07(2014)079

[211] Dr. D. Dercks, Private communication: Dr. Dercks performed the MadGraph calculation in the context of a CheckMATE analysis.

[212] H. Baer, C. H. Chen, F. Paige, X. Tata, "Trileptons from Chargino-Neutralino Production at the CERN Large Hadron Collider," *Phys.Rev. D*, vol. 50, pp. 4508–4516, 1994, arXiv:hep-ph/9404212v1. doi:10.1103/PhysRevD.50.4508

[213] Dr. D. Dercks, Private communication: Dr. Dercks performed the CheckMATE analysis.

[214] M. Drees, H. Dreiner, D. Schmeier et al., "CheckMATE: Confronting your Favourite New Physics Model with LHC Data," *Comput.Phys.Commun.* , vol. 187, pp. 227–265, 2015, arXiv:1312.2591 [hep-ph]. doi:10.1016/j.cpc.2014.10.018

[215] D. Dercks, N. Desai, J.S. Kim et al., "CheckMATE 2: From the model to the limit," *Comput.Phys.Commun.*, vol. 221, pp. 383–418, 2017, arXiv:1611.09856 [hep-ph]. doi:10.1016/j.cpc.2017.08.021

[216] The ATLAS Collaboration, "Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s}$=13 TeV using the ATLAS detector," *Phys. Rev. D*, vol. 94, p. 032005, 2016, arXiv:1604.07773v2 [hep-ex]. doi:10.1103/PhysRevD.94.032005

[217] The ATLAS Collaboration, "Search for squarks and gluinos in final states with jets and missing transverse momentum at $\sqrt{s}$=13 TeV with the ATLAS detector," *Eur. Phys. J. C*, vol. 76, p. 392, 2016, arXiv:1605.03814v2 [hep-ex]. doi:10.1140/epjc/s10052-016-4184-8

[218] The ATLAS Collaboration, "Search for squarks and gluinos in events with an isolated lepton, jets and missing transverse momentum at $\sqrt{s}$=13 TeV with the ATLAS detector," 2016, contribution to 38th International Conference on High Energy Physics. https://cds.cern.ch/record/2206136/files/ATLAS-CONF-2016-054.pdf

[219] The ATLAS Collaboration, "Search for direct top squark pair production and Dark Matter production in final states with two leptons in $\sqrt{s}$=13 TeV pp collisions using 13 fb^{-1} of ATLAS data ," 2016, ATLAS-CONF-2016-076. http://cds.cern.ch/record/2206249/files/ATLAS-CONF-2016-076.pdf

[220] The ATLAS Collaboration, "Search for squarks and gluinos in final states with jets and missing transverse momentum using 36 fb^{-1} of $\sqrt{s}$=13 TeV pp collision data with the ATLAS detector," 2017, presented at 52nd Rencontres de Moriond on Electroweak Interactions and Unified Theories, La Thuile, Italy, 18 - 25 Mar 2017. http://cds.cern.ch/record/2258145/files/ATLAS-CONF-2017-022.pdf

[221] The ATLAS Collaboration, "Search for an invisibly decaying Higgs boson or dark matter candidates produced in association with a Z boson in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector ," presented at the 2017 European Physical Society Conference on High Energy Physics, Venice, Italy, 05 - 12 Jul 2017. http://cds.cern.ch/record/2273610/files/ATLAS-CONF-2017-040.pdf

[222] D.P. Bertsekas, *Nonlinear Programming.* Athena Scientific, 1999.

[223] A. K.E., *An introduction to numerical analysis.* Wiley, 1978.

[224] B. List, J. List, "An Object-Oriented Kinematic Fitting Package," 2009, LC-TOOL-2009-001.

[225] R.J. Barlow, *Statistics - A Guide to the Use of Statistical Methods in the Physical Sciences.* Wiley & Sons, 1999.

[226] J. Tian, "Study of Higgs self-coupling at the ILC based on the full detector simulation at $\sqrt{s}$=500 GeV and $\sqrt{s}$=1 TeV ," vol. LC-REP-2013-003, 2013.

[227] J. Tian - Private communication, 2015.

[228] E. Farhi, "Phys. Rev. Lett." vol. 39, pp. 1587–1588, 1977.

[229] Ian Brock, Thomas Schörner-Sadenius, *Physics at the Terascale.* John Wiley & Sons, 2011.

[230] B. A. Shenoi, *Introduction to Digital Signal Processing and Filter Design.* John Wiley & Sons, 2005.

[231] R. Hartley, A. Zisserman, *Multiple View Geometry in computer vision.* Cambridge Press, 2000.

[232] D. Forsyth, J. Ponce, *Computer Vision: A Modern Approach.* Prentice Hall, 2003.

[233] J.M.S. Prewitt, "A 3x3 Isotropic Gradient Operator for Image Processing," pp. 271–272, 1973.

[234] J.M.S. Prewitt, "Object Enhancement and Extraction," pp. 75–149, 1970.

[235] J. Canny, "A Computational Approach to Edge Detection," vol. PAMI-8, pp. 679–698, 1986.

[236] S. Caiazza - Private communication.

[237] T. Suehara, J. List, "Chargino and Neutralino Separation with the ILD Experiment," Internal Note. https://confluence.desy.de/download/attachments/57088733/poin t5_090319_2.pdf?version=1&modificationDate=1510654435828&api=v2

[238] E. Davies, *Machine Vision: Theory, Algorithms and Practicalities.* Academic Press, 1990.

[239] F. James, "A review of pseudorandom number generators," *Comp. Phys. Comm.*, vol. 60, pp. 329–344, 1990. doi:10.1016/0010-4655(90)90032-V

[240] M. Lüscher, "A portable high-quality random number generator for lattice field theory simulations," *Comp. Phys. Comm.*, vol. 79, pp. 100–110, 1994. doi:10.1016/0010-4655(94)90232-1