University of Warsaw
Faculty of Physics

Jan Franciszek Klamka
Student's book number: 373299

Searching for Inert Doublet Model scalars at high energy CLIC

Second cycle degree thesis
Field of study: Physics

CERN-THESIS-2020-098
01/07/2020

The thesis has been written under the supervision of
prof. dr hab. Aleksander Filip Żarnecki
Particles and Fundamental Interactions Division
Institute of Experimental Physics

Warsaw, July 2020

Abstract

The Inert Doublet Model (IDM) is a simple extension of the Standard Model, introducing an additional Higgs doublet that brings in four new scalar particles. The lightest of the IDM scalars is stable and is a good candidate for a dark matter particle. The potential of discovering the IDM scalars in the experiment at Compact Linear Collider (CLIC), e^+e^- collider proposed as the next generation infrastructure at CERN, has been tested for two high-energy running stages, at 1.5 TeV and 3 TeV. Considered is the CLIC sensitivity to IDM charged scalar pair-production for 23 high-mass benchmark points in the model parameter space and the semi-leptonic final state. The simplified simulation of detector response with Delphes package was used, and the results were validated for the selected 5 benchmark scenarios using full detector simulation based on Geant4. For both simulation methods the cut-based event preselection was applied, followed by a multivariate analysis based on the Boosted Decision Trees. Results of the study indicate that heavy charged IDM scalars can be discovered at CLIC for most of the proposed benchmark scenarios, with statistical significance, depending on the model parameters, up to about 60σ.

Keywords

Compact Linear Collider, Inert Doublet Model, scalar particles, dark matter, computer simulations, multivariate data analysis, machine learning, boosted decision trees

The title of the thesis in Polish

Poszukiwanie cząstek skalarnych modelu IDM przy akceleratorze CLIC

Contents

1. Introduction 3

2. The Compact Linear Collider 5
- 2.1. Accelerator 5
- 2.2. Experiment 6
- 2.3. Analysis tools 7

3. The Inert Doublet Model 11
- 3.1. Motivation 11
- 3.2. The Model 12
- 3.3. Benchmark points 13
- 3.4. Search strategies 14
- 3.5. Limitations of the leptonic channel searches 15

4. Data processing 19
- 4.1. Signal and background samples 19
- 4.2. Fast simulation 19
- 4.3. Full simulation 21

5. Fast simulation analysis 23
- 5.1. Event preselection 23
- 5.2. Multivariate analysis 27
- 5.3. Results 28

6. Full simulation analysis 31
- 6.1. Event preselection 31
- 6.2. Multivariate analysis 35
- 6.3. Results 35
- 6.4. Validation of fast simulation 36

7. Conclusions 39

Appendix 41

Chapter 1

Introduction

While the Standard Model (SM) is considered a well-tested and well-established theory of the elementary particle physics, there are many indications, that it is not the final theory describing the Universe. The phenomena of dark matter and dark energy, barion assymetry, hierarchy and naturalness problems remain unexplained, and these are only the main examples. However, no signs of physics Beyond Standard Model (BSM) have yet been seen at the Large Hadron Collider (LHC) and the Higgs boson discovery only confirmed predictions of the SM. To extend tests of SM, many new collider projects are being developed, promising new ways of searching the "new physics" phenomena. Among these, many lepton collider options are proposed, as they would allow very precise measurements and fine tests of the SM predictions (and predictions of different established BSM theories) using approach which is complementary to LHC searches for the BSM physics. The Compact Linear Collider is one of the e^+e^- collider projects, where electrons could be collided with positrons at the energy scales from few hundred GeV up to the multi-TeV, starting around 2035.

Before the final decision on the collider project can be reached, its potential has to be thoroughly tested. The decision has to be based on the comparison of different projects, and one needs to find out, which one is the most beneficial (from the scientific point of view) and affordable at the same time, and should thus be implemented in the future. Therefore, detailed studies with the use of realistic simulations are performed, to show the advantages and physical potential of different collider projects. In this way, the potential sensitivity of the future experiments to particular BSM scenarios can be tested.

Theoretical physicists around the globe come up with countless models to solve the problems of Standard Model, most commonly proposing the existence of some new, unknown particles. These theories have to be experimentally tested to check, whether they truly describe the nature or not. This is also the case for the Inert Doublet Model (IDM), which is a very simple and straightforward extension of the SM, being able to explain many of its problems, like the phenomenon of dark matter. It brings in four new scalar particles, one of which is a natural dark matter candidate.

The aim of this work was to investigate the Compact Linear Collider potential for discovering IDM scalars. It was previously studied for a channel, where the pair-production of these hypothetical particles resulted in the signature of two leptons in the final state; however, the statistics in this channel turned out to be too low for the discovery of high mass scenarios. To increase statistical significance of observations, considered in this thesis is the search of IDM scalar pair production in the semi-leptonic channel, with the two jets, lepton and neutrino in the final state.

Information on the Compact Linear Collider project, including the concepts of accelerator and experiment (detector design), as well as the software tools used to perform the study, are described in Chapter 2. The physical motivations for the introduction of the Inert Doublet Model, the model formulation and details on the benchmark scenarios, as well as description of the search strategy and results of the previous analysis based on leptonic final state can all be found in Chapter 3. In Chapter 4 detailed information about the samples used for the study is given, as well as details on the event reconstruction, for both the fast and full detector simulation analyses. Description of the event pre-selection and selection, together with all the results, can be found in Chapters 5 and 6, for fast and full simulation studies, respectively.

All results presented in this thesis, starting from Chapter 4, were obtained by the Author, who was also responsible for running the fast detector response simulations, event reconstruction and preparing the analysis framework for both fast and full simulation analysis flow. This study was carried out within the OPUS project, supported by the National Science Centre, Poland (nr 2017/25/B/ST2/00496). Preliminary results of this work were presented by the group members at the international conferences [1, 2].

Chapter 2

The Compact Linear Collider

The Compact Linear Collider (CLIC) is a concept of high-luminosity, multi-TeV linear e^+e^- collider. It is being developed by the CLIC collaboration, that in 2012 published Conceptual Design Report (CDR) for CLIC [3]. Responsible for the studies of CLIC physics potential, the research and development (R&D) on the detector technologies, and for the detector design is the CLIC Detector and Physics (CLICdp) collaboration, which was formed as well in 2012 [4]. University of Warsaw is an active member of CLICdp since 2015.

CLIC and CLICdp collaborations submitted a series of documents as an input the update of the European Strategy for Particle Physics (a document describing prioritisation of research on particle physics in Europe) which have been just approved by the CERN Council.[1] These documents summarize all the studies made within CLIC and CLICdp in the recent years [5].

2.1. Accelerator

CLIC accelerator is the only mature proposal of the multi-TeV lepton collider, based on a novel two-beam normal-conducting accelerating technology, that allows to reach high accelerating gradients of the order of 100 MV/m. A scheme of the accelerator is presented in Figure 2.1. The high-current, low-energy drive beam is used to generate many short and strong radio-frequency pulses, which are then distributed alongside the main linac to accelerate the low-current main beam to the high energies. This approach results in higher efficiency and lower costs, compared to accelerating the beam directly with klystrons [7].

[1] As the Strategy was only approved on June 19, when the thesis was almost finished, recommendations of the Strategy are not reflected in the text.

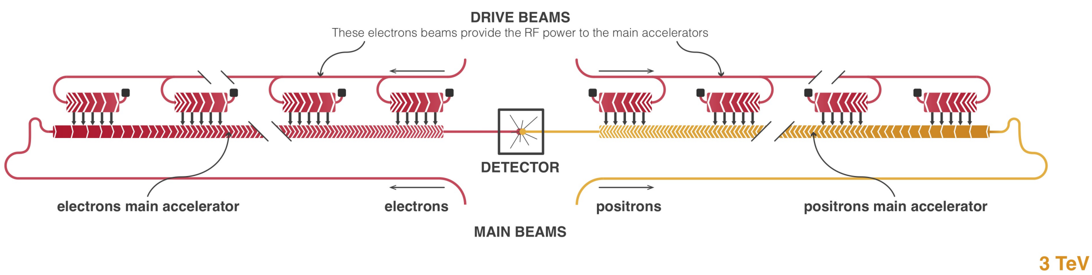

Figure 2.1: Layout of the CLIC complex at the 3 TeV energy stage [6].

The accelerator is planned to be built at the CERN site. Assuming the construction starts around 2026, one could expect the first beam collisions around 2035 [8]. Three centre-of-mass energy stages are assumed. The first one, at 380 GeV, should allow for the top quark pair-production (including the threshold scan), as well as for detailed studies of the Higgs boson couplings in the Higgstrahlung and the WW-fusion processes. Subsequent CLIC upgrades to 1.5 TeV and 3 TeV are planned with 7 to 8 years of data taking at each energy stage. Corresponding total integrated luminosities are $1\,\mathrm{ab}^{-1}$, $2.5\,\mathrm{ab}^{-1}$ and $5\,\mathrm{ab}^{-1}$ for 380 GeV, 1.5 TeV and 3 TeV respectively. The proposed CLIC running schedule offers the high energy reach, as well as the possibility of many precise measurements, based on large statistics of collected data, with the total programme duration of 25-30 years. In the baseline design, only $\pm80\%$ polarisation of the electron beam is possible (with no polarisation of the positron beam). At the first energy stage, the running time is to be split equally between negative and positive beam polarisation, while at the higher energy stages, the 80:20 luminosity sharing is assumed, in favour of -80% polarisation. This choice is a compromise between the Higgs boson production through the WW boson fusion (which increases with -80% electron polarisation) and the full sensitivity for the Beyond Standard Model (BSM) effects (which requires running also with +80% beam polarisation) [4].

2.2. Experiment

For the considered CLIC experiment, a dedicated detector model was designed and optimised for the concept of Particle Flow analysis (explained below). Schematic view of the CLIC detector (CLICdet) model, with a description of its components, is presented in Figure 2.2. The detector has a shape of a barrel and a layered structure, typical for the high energy physics collider experiments. The interaction point in the centre of the detector is surrounded by a silicon pixel vertex detector and a silicon tracker. The next layers are (ordered from the inside): electromagnetic and hadronic calorimeters (ECAL and HCAL), superconducting 4 T solenoid, and the iron yoke, instrumented with muon chambers [10].

Precision measurements considered at CLIC require the best possible reconstruction of all final state objects, including the hadronic jets. Proper reconstruction of the hadronic invariant mass is crucial for many precision studies (eg. hadronic top-quark decays) as well as for many BSM searches. The main objective of the detector design was to achieve jet energy resolution allowing to distinguish between W and Z boson hadronic decays through the invariant masses of their decay products. This requires the jet energy resolution of the order of $\sigma_E/E \lesssim 3.5\% - 5\%$ [11]. Jet energy measurement in calorimeters, however, is limited by the fluctuations in the hadronic cascade development. Nevertheless, it can be improved by the approach called Particle Flow Calorimetry. All particles contained in the jet should be first reconstructed one-by-one. Energy of the neutral particles has to be measured in the calorimeters, but for the charged ones (which are the majority) a much more precise momentum measurement in the tracking system can be used. If one can properly match the tracks to the deposits in calorimeters (to remove clusters originating from the charged particles), the resolution of the jet energy (calculated as the sum of neutral and charged particle energies) can be significantly improved [12].

To profit from the Particle Flow approach, fine-grained calorimeters are required, so the deposits of the charged particles can be accurately matched with tracks in the central tracking system. The tracks themselves need to be measured very precisely, in a wide angular and energy range, to obtain accurate estimates of their momentum, and thus, the energy. Simulation studies show, that with the resolution of the inverse transverse momentum, $1/p_T$,

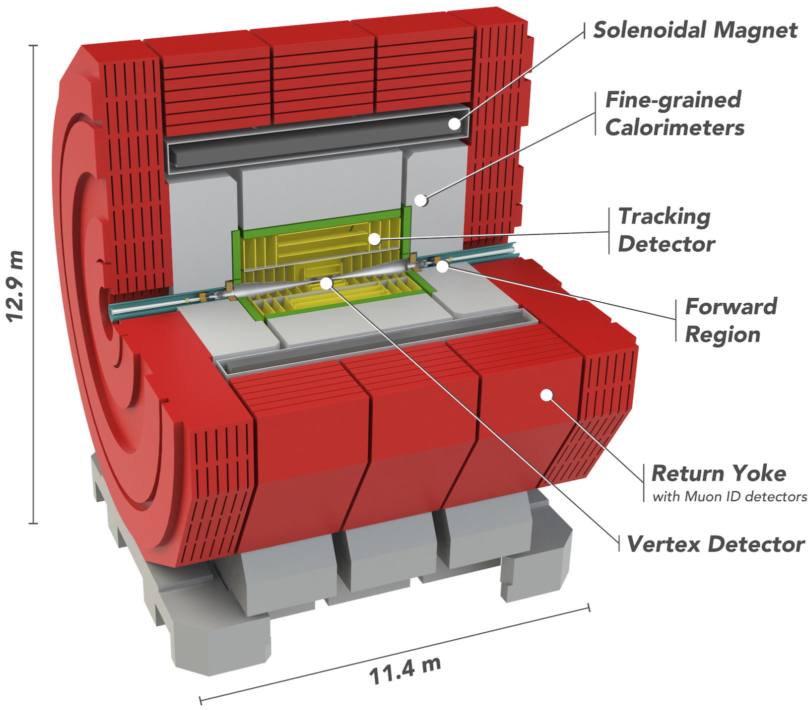

Figure 2.2: 3D scheme of the CLICdet [9].

reaching $2 \cdot 10^{-5}\,\mathrm{GeV}^{-1}$ at large track angles, the jet energy resolution achievable at CLIC is about 3.5-4%, for high-energy jets. Furthermore, the silicon pixel vertex detector placed close to the beam provides the impact parameter resolution $\sigma_d < 5\,\mu$m for tracks with high momentum, which should allow for very efficient flavour tagging [13].

2.3. Analysis tools

To fully analyze the physics potential of experiments at future colliders, including e^+e^- colliders such as CLIC, a strong software background is needed so the realistic predictions can be made.

The first tool is necessary for the computation of the cross sections, decay products and kinematic distributions of the observables in physical processes considered. In case of the e^+e^- collider projects this is usually provided by the WHIZARD package [14], which also allows to generate the event samples. The matrix elements at the leading order in perturbation theory are automatically calculable and other features may be added as well (such as higher orders). One can also internally call PYTHIA [15] to simulate parton showering and hadronization.

The next step is to simulate the expected detector response to generated, incoming particles. To make it fully realistic, one has to simulate the passage of particles through matter. This is possible with the GEANT4 package [16], which includes particle tracking, detector geometry description, different physics models and simulation of detector hits (and, like WHIZARD, uses Monte Carlo methods). However, these simulations require high computing power. They can be performed on machines of everyday use, but still the calculations take a lot of time on them. Therefore, the most common way of running the detector response simulations is to use the *Worldwide LHC Computing Grid* [17] for this purpose. It is a global computer network, that was created to provide computing and data storage resources for handling and analysing huge amounts of data from LHC, but is also used by other projects.

Because of time and effort needed for the simulation of detector response, a concept of the so-called fast simulation came up (in contrast to the previously mentioned *full* simulation), which is a simplified approach to the problem. An example of tool introducing the fast simulation is DELPHES package [18]. In this framework, the track propagation in a magnetic field is included, as well as inside ECAL, HCAL and muon identification system. The detector response is parametrised as a function of the particle type, energy and direction, taking into account detector geometry, granularity of calorimeters, sub-detector resolutions and the effect of the magnetic field. What the fast simulation does neglect, is the exact simulation of the particle propagation through matter and resulting energy losses.

DELPHES was initially intended as the simulation tool for hadron colliders, but yet, in an easy way, it can also be used for modeling the e^+e^- experments. In particular, CLICdet model cards (a DELPHES configuration file for CLICdet) were implemented to allow fast simulation studies for CLICdp [19].

The beam-beam interactions resulting in beamstrahlung and the radiative energy losses of the incoming beams are the main reasons for the energy of colliding electrons and positrons to be different from the nominal beam energy. The influence of heterogeneous beam luminosity spectrum can be significant – in particular at high CLIC running energies – and has to be taken into account in the simulation studies. The expected beam energy spectra[2] for CLIC running at 380 GeV and 3 TeV are shown in Figure 2.3.

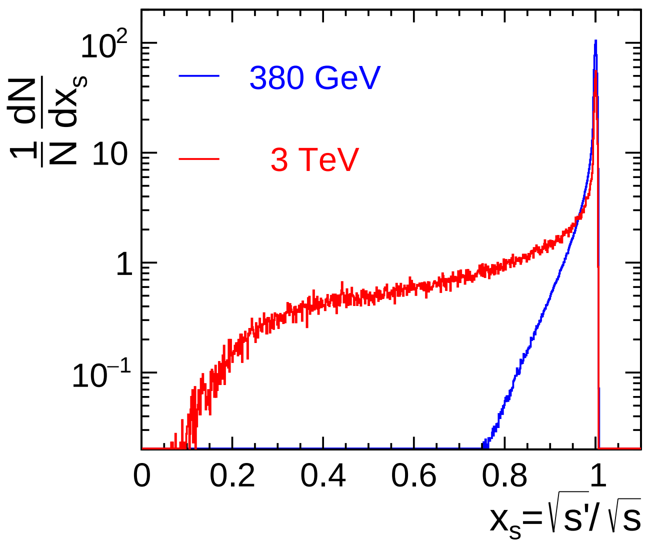

Figure 2.3: The luminosity spectra for CLIC operating at $\sqrt{s}$ = 380 GeV and 3 TeV, where x_s denotes the ratio of the e^+e^- centre-of-mass energy after beamstrahlung, to the nominal centre-of-mass energy [3].

[2]The effects of the initial state radiation (ISR), the photon emission by $e^{\pm}$ just before the hard collision, are not included in the beam spectra as they are process dependent. They are included in the process simulation withing the WHIZARD framework.

Multivariate analysis tools

A standard way to deal with the signal extraction problem is the cut-based event selection. Subsequent cuts are used to reduce background contamination of the signal sample. In recent years, however, techniques incorporating *machine learning* algorithms became widely used and turned out to be very efficient in this type of applications. Machine learning (ML) is an application of statistical methods and high computing power to estimate complicated functions [20], especially the multidimensional ones (including probability density functions, PDFs). ML algorithms are designed to be able to optimise, improve their performance (learn) by *training* – running repeatedly on the so-called training dataset, comparing the result to the expectations (ground truth) and, based on the analysis of the differences, adjusting the algorithm's parameters. In high energy physics, one of the common objectives is to assign events to some particular classes, based on their topology and kinematic properties. The events are characterized by values of many variables, where a distribution of each variable in a set of events might be described by a PDF. If one knew such multivariate PDFs (including possible correlations between individual variables) for each considered class of events, it would be much easier, and more *probable*, to classify an event to the correct class, given its kinematic properties. High energy physics is therefore a perfect example of field where ML might be very useful, providing a way to estimate those multidimensional functions. In short, it gives a possibility to extend the dimensionality of analysis, as human perception is limited only to analysis in two, or maximally 3 dimensions.

Boosted Decision Trees

Many instances of ML algorithms exist, with neural networks being probably the most famous one. In this thesis, an algorithm taking advantage of the so-called Boosted Decision Trees (BDT) was used, implemented in TMVA Toolkit [21] integrated into the ROOT analysis framework [22].

A decision tree is a binary classifier, in which decisions (cuts) divide the parameter space (in this case: kinematic variables space) multiple times. Each cut is a "branch", and resulting parts of space are "leafs", which are then further cut into smaller ones. After all cuts are made, algorithm calculates the probability for each of the final (output) leafs to belong to the given class – depending on this score all events in the space region of the leaf are assigned to particular class.

However, both cuts and scores assignment, in a regular decision tree, are completely random – it never gives the reasonable results at first. One has to prepare an algorithm, that will "learn" on previous choices – this is called *boosting* the decision trees. In this approach, a set of trees (forest) is used. Each tree subsequently classifies the events, and in the end gets a score, depending on how close it was to the ground truth. The trees are then combined into the single, powerful classifier, in which the influence of the results from particular decision trees depend on the score they have received. This allows to significantly improve the effectiveness of decision trees.

Chapter 3

The Inert Doublet Model

3.1. Motivation

Formulation and verification of the Standard Model (SM) was one of the greatest achievements in physics in the 20th century, crowned with the discovery of the Higgs boson in 2012 by ATLAS and CMS experiments at CERN [23, 24]. SM exploits Quantu Field Theory (QFT) approach to describe all known elementary particles and interactions between them: electromagnetic, weak and strong interactions (gravity is not yet understood on QFT level). Fundamental SM particles are: three families of quarks, leptons and neutrinos (fermions), gauge vector bosons mediating interactions (electromagnetic: photon, strong: gluons, weak: Z^0 and $W^{\pm}$ bosons), and the scalar Higgs boson. The SM Lagrangian is symmetric under $SU(3) \times SU(2) \times U(1)$ group, and the key elements of the SM are: the electroweak $SU(2) \times U(1)$ spontaneous symmetry breaking (EWSB) [25] and the Brout-Englert-Higgs mechanism [26]. Although numerous experiments confirmed the vast majority of model assumptions (especially, all the results from Large Electron-Positron Collider, LEP, and LHC are properly described), still several puzzles and clues imply, that SM is not the whole picture.

First, and probably most famous problem, is the discrepancy between contents of the Universe expected from the SM predictions and inferred from the astrophysical and cosmological observations. Cosmic microwave background (CMB) measurements, with the latest coming from the Planck satellite mission [27], show that that the Universe has no spatial curvature. However the density of matter in the Universe is only about 31% of the so-called critical density, which one would expect for the flat universe. The other 69% of critical density – widely called *dark energy* – is attributed to the Λ constant present in the Einstein equation [28]. From 31% of critical density, baryonic matter described by the SM makes up only about 5%, as indicated by the Planck analysis. This suggests that the remaining 26% correspond to some new form of matter, called *dark matter* (DM), that is not baryonic, non-luminescent, cold, and not described by the SM. There are also other hints suggesting the existence of DM, like the observations of the rotation curves of galaxies, or in the gravitational lensing.

Another problem of the SM is the so-called baryon asymmetry, which is the difference between amount of matter and antimatter observed in the Universe. In the early Universe, just after the Big Bang, equal amounts of matter and antimatter should be produced due to the charge symmetries (and also no reasonable cause for the different amounts to be produced). Andrei Sakharov proposed three conditions, which have to be fulfilled for this asymmetry to appear [29]. First: baryon number has to be violated, second: charge (C) and charge-parity (CP) symmetries have to be violated, and third: at some point in the Universe expansion there had to be no thermal equilibrium. The first condition is not ensured by the SM at all and CP violation expected in the SM is also insufficient for the second condition to be fulfilled (the third one is not relevant as SM does not describe the evolution of the Universe).

Finally, many questions and problems remain unsolved in the SM. For instance, there is no particular reason for the matter to be made of fermions and for the interactions to be mediated by bosons – SM provides no explanation to this. Also, neutrinos in the SM are massless, even though it was already experimentally proven that they indeed have masses [30]. The reason for EWSB to occur is also not explained in the SM.

When considering possible extensions of the SM it is important to notice that scalar potential in the SM Lagrangian has not been tested experimentally, yet. Only the mass of the Higgs boson was established, but this is not enough to test validity of the model. The Higgs self-coupling has to be measured to confirm the shape of the scalar potential, but it will be a very challenging task not only for the LHC and HL-LHC, but also for future colliders.

These are the main reasons for which people do believe that the SM is not the final "theory of everything" but rather an approximate theory, which still has to be extended. That is why physicists search for the BSM effects and new particles, and the model considered in this thesis is one of the proposed SM extensions to be tested at future colliders.

3.2. The Model

The Inert Doublet Model was introduced in 1977 [31]. It is one of the simplest and most straightforward extensions of the SM, providing an explanation of the dark matter phenomenon. In the description of the model convention and assumptions introduced in [32] were adopted.

The scalar potential in this model can be written as

$$V = -\tfrac{1}{2}\left[m_{11}^2\left(\phi_S^\dagger\phi_S\right) + m_{22}^2\left(\phi_D^\dagger\phi_D\right)\right] + \tfrac{\lambda_1}{2}\left(\phi_S^\dagger\phi_S\right)^2 + \tfrac{\lambda_2}{2}\left(\phi_D^\dagger\phi_D\right)^2 + \lambda_3\left(\phi_S^\dagger\phi_S\right)\left(\phi_D^\dagger\phi_D\right)$$
$$+\lambda_4\left(\phi_S^\dagger\phi_D\right)\left(\phi_D^\dagger\phi_S\right) + \tfrac{\lambda_5}{2}\left[\left(\phi_S^\dagger\phi_D\right)^2 + \left(\phi_D^\dagger\phi_S\right)^2\right], \tag{3.1}$$

where ϕ_S and ϕ_D are the two SU(2) doublets of complex scalar fields, with ground states $\langle\phi_S\rangle = (0, v/\sqrt{2})$ and $\langle\phi_D\rangle = 0$ and v being vacuum expectation value (vev). The potential is symmetric under the Z_2 group, with the doublets transforming as follows: $\phi_S \to \phi_S$, $\phi_D \to -\phi_D$. The decomposition around the vacuum state is given by:

$$\phi_S = \begin{pmatrix} \phi^+ \\ \frac{1}{\sqrt{2}}(v + h + i\xi) \end{pmatrix}, \quad \phi_D = \begin{pmatrix} H^+ \\ \frac{1}{\sqrt{2}}(H + iA) \end{pmatrix}. \tag{3.2}$$

With the assumed ground states, ϕ_S is the SM-like Higgs doublet involved in the EWSB, h is the Higgs field, and $\phi^+ \in \mathbb{C}$ and ξ correspond to the three Goldstone bosons responsible for masses of $W^\pm$ and Z bosons. The second doublet ϕ_D is called inert (or dark); it contains new scalar fields $H^\pm, H, A$ and does not participate in the EWSB.

The Z_2 symmetry, under which all the SM fields are even, preserves the dark scalars from coupling to SM fermions on tree-level. Being discrete, it also results in the conservation of a new, multiplicative quantum number, so the lightest of new scalars must be stable – and hence, it is a natural dark matter candidate.

Based on the results from LHC, SM-like Higgs boson mass and vev can be fixed to $m_h \equiv$

$m_{11} = \sqrt{\lambda_1}v^2 \simeq 125$ GeV and $v \simeq 246$ GeV. Masses of the four new particles are given by

$$\begin{aligned} m_{H^\pm}^2 &= \frac{1}{2}\left(\lambda_3 v^2 - m_{22}^2\right), \\ m_A^2 &= m_{H^\pm}^2 + \frac{1}{2}\left(\lambda_4 - \lambda_5\right) v^2, \\ m_H^2 &= m_{H^\pm}^2 + \frac{1}{2}\left(\lambda_4 + \lambda_5\right) v^2, \end{aligned} \tag{3.3}$$

where either H or A is the lightest (as the charged dark matter is excluded). Here H will be considered as the lightest inert scalar (DM particle), which implicates $\lambda_4 + \lambda_5 < 0$ and $\lambda_5 < 0$. There is a freedom of choice here, because $\lambda_5 > 0$ would only results in A being the lightest particle. The two scenarios (H or A selected as DM particle) are equivalent: the CP transformation is not relevant to the dark scalars, since H and A do not couple to fermions.

It is also useful to define $\lambda_{345} = \lambda_3 + \lambda_4 + \lambda_5$, as it binds couplings related to the triple and quartic coupling of h to H. After the EWSB and fixing m_h and v to the SM values, the model can be described by 5 free parameters. As in [32], the following set of physical parameters is selected: three scalar masses, $m_{H^\pm}$, m_A and m_H, and two couplings, λ_2 and λ_{345}. For more information about the model, please refer to [32, 33] and references therein.

3.3. Benchmark points

To simulate the production of IDM particles in the experiment, one has to select benchmark points with fixed values of the model parameters – points in the 5-D space spanned by the free parameters. These were selected in [33] from the scan over the parameter space. The points were chosen to cover all interesting areas of space and to respect widely diverse features of production channels considerable at lepton colliders. They were also assumed to be accessible at different collision energies: from 250 GeV at the International Linear Collider (ILC) and 380 GeV at the first stage of CLIC, up to multi-TeV scales at the high energy stages of CLIC. Scalar masses for the considered points from the parameter space scan and for the benchmark points selected for further analysis are presented in Figure 3.1.

The benchmark points shown in Figure 3.1 respect all the theoretical, and the current experimental constraints. Considered constraints come from: Higgs boson mass, width and signal strenghts at LHC, gauge bosons widths, electroweak precision measurements, boundary estimates of the charged dark scalar mass and lifetime, searches for new physics in the LHC, LEP and Tevatron, and dedicated dark matter searches or phenomenology measurements, e.g. from Planck [27], Fermi-LAT [34] or latest results from XENON1T experiments [35]. It is interesting to notice that, when XENON1T finishes collecting the data, only four of the benchmarks will be out of its reach.

Considered in this thesis are 23 out of 41 benchmark points presented in [33]. As the study concerns high energy stages of CLIC, all high mass benchmark points (HP) and three low mass scenarios (BP) with high scalar masses: BP18, BP21 and BP23 were selected. The parameter values of these points can be found in Table A.1 in Appendix, together with the expected DM relic density in the Universe, Ω_H, scaled by the Hubble constant squared, h^2 (where h is given in units of $100\,\mathrm{km/s/Mpc}$). More information, as well as exact constraints on the model, can be found in [33].

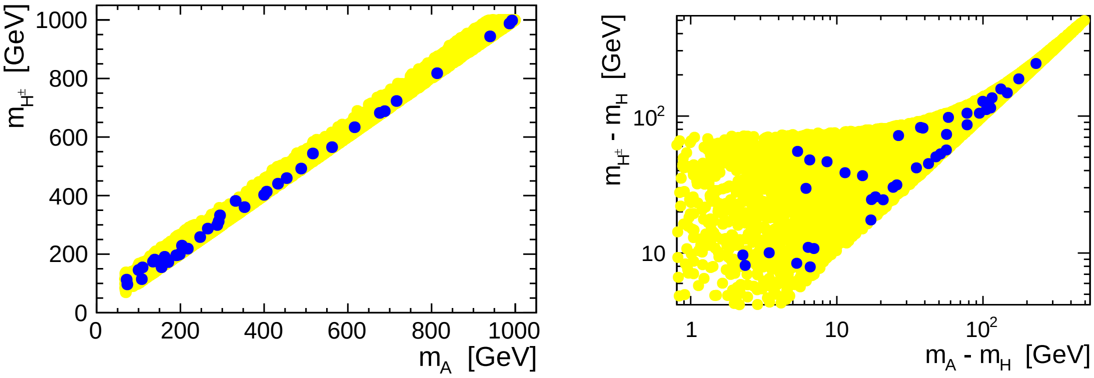

Figure 3.1: Benchmark points presented on the planes of scalars' masses (left) and their differences (right) [33]. Yellow dots represent points from the parameter scan consistent with all theoretical and experimental constraints, and the blue dots are the selected benchmark points.

3.4. Search strategies

Inert scalars can only be pair-produced via their couplings to the SM gauge bosons. Detection of the IDM scalars at lepton colliders can be considered in the two production channels, neutral or charged scalar pair-production:

$$
\begin{aligned}
e^+e^- &\rightarrow H\ A\,,\\
e^+e^- &\rightarrow H^+H^-.
\end{aligned}
$$

The leading-order cross sections for these processes are presented in Figure 3.2 for collision energies of 1.5 TeV and 3 TeV considered in this thesis. The cross sections are determined by the scalar masses, the dependence on the couplings λ_2 and λ_{345} is marginal.

As the A scalar always decays to $Z\,H$ and the dominant decay channel for charged inert scalars is $W^\pm\ H$, the pair-production of DM particles (H) can be searched for in two channels:

$$
\begin{aligned}
e^+e^- &\rightarrow H\ A \rightarrow Z\ H\ H\,,\\
e^+e^- &\rightarrow H^+H^- \rightarrow W^+W^-H\ H\,. \qquad (3.4)
\end{aligned}
$$

Different experimental signatures can be considered depending on the decay modes of the Z and $W^\pm$ bosons. Shown in Figure 3.3 are the Feynman diagrams corresponding to these processes. Diagram for the leptonic channel is shown for the neutral scalar pair-production and the semi-leptonic channel diagram for the charged scalar pair-production, when one of the W bosons decays into lepton and neutrino, and another one into two quarks. Note, that the leptonic final state is possible in both production channels (when both W bosons decay into the lepton-neutrino pair and the two leptons can be detected), while the semi-leptonic is only possible for the charged scalar pair-production. It is also worth noting, that Z and W$^\pm$ bosons decaying into the final state fermions may be both on- and off-shell, depending on the mass differences, $m_A - m_H$ and $m_{H^\pm} - m_H$.

Expected signatures of the inert scalar pair-production are isolated leptons or lepton pairs, or jets, and a large missing energy-momentum, as DM particles (dark scalars H) do not interact with matter and escape from the detector. A lepton pair of the same flavour (for

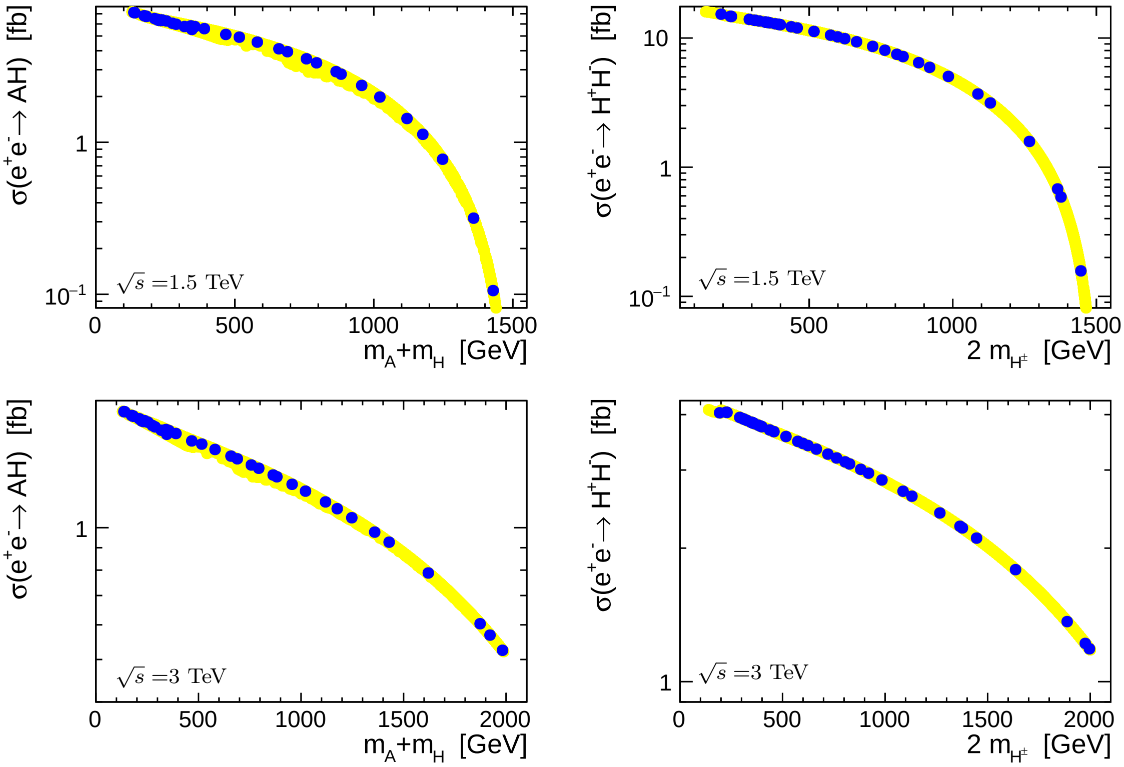

Figure 3.2: Leading-order cross sections for neutral (left) and charged (right) inert scalar production, $e^+e^- \to A\ H$ and $e^+e^- \to H^+H^-$, for collision energy of 1.5 TeV (upper plots) and 3 TeV (lower plots). The yellow band represents all scenarios selected in the model scan [33] while the blue dots represent the selected benchmark scenarios. Beam energy spectra are not included [2].

both HA and H^+H^- production channels), or with different flavours (only possible in H^+H^- pair-production) can be considered for the leptonic channel. It provides the cleanest signal, as isolated leptons (especially muons) are the most efficiently reconstructed particles in collider experiments. However, colour charge of quarks allows $W^\pm$ boson to decay into hadrons 2 times more frequently, than to leptons. Therefore, semi-leptonic channel (with electron or muon in the final state and a pair of jets) is expected to yield almost an order of magnitude higher statistics for H^+H^- production signal than different flavour lepton pair signature.

3.5. Limitations of the leptonic channel searches

Search for the IDM particles at CLIC was previously studied for the leptonic channel, results are described in [36, 37]. The study was based on event samples generated with Whizard 2.2.8, taking into account the CLIC beam energy spectra, but no detector response simulation was made – only cuts reflecting detector acceptance were applied. Collisions of unpolarised beams were simulated, assuming $2.5\,\mathrm{ab}^{-1}$ and $5\,\mathrm{ab}^{-1}$ luminosity for the 1.5 TeV and 3 TeV energy stages respectively. Both production channels possible at lepton colliders were considered, with assumed signatures: $\mu^+\mu^-$ for the HA channel (to obtain maximum purity), and $e^\pm\mu^\mp$ for the H^+H^- production. Cut-based pre-selection, followed by a multivariate analysis (MVA), were used to calculate the expected signal significance for different IDM benchmark

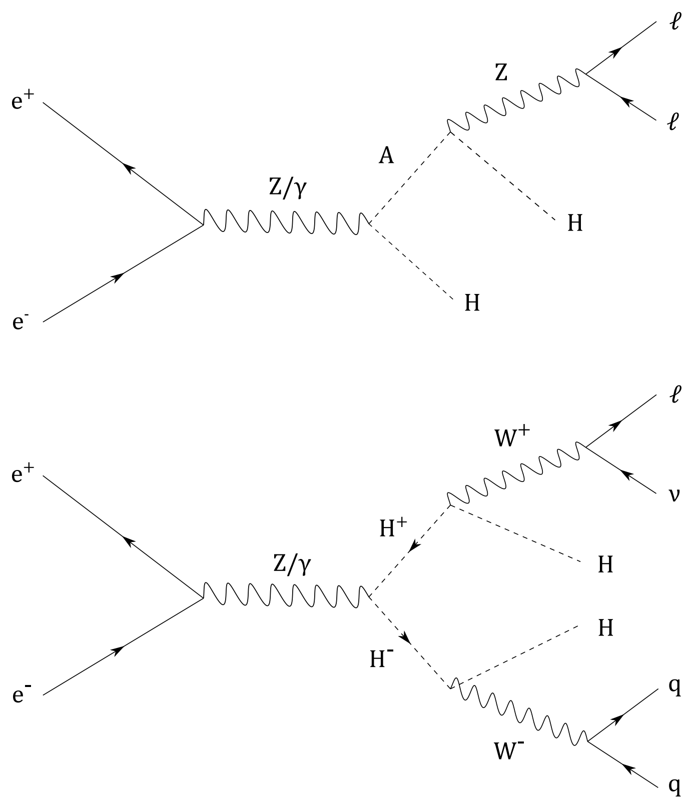

Figure 3.3: The leading order Feynman diagrams for the possible production channels of the IDM particles in lepton colliders. The upper one shows neutral scalar pair-production with the fully-leptonic final state, and the lower one shows the charged scalar pair-production with the semi-leptonic final state.

points and different CLIC stages.

Results of the study [36] are summarised in Figure 3.4. Statistical significance of the deviations from the SM predictions, expected for considered benchmark point, is plotted as a function of sum of the masses of produced IDM particles ($m_A + m_H$ for $e^+e^- \to ZA$ and $2\,m_{H^\pm}$ for $e^+e^- \to H^+H^-$), for three considered CLIC stages. Significance describes the difference between the result expected for the signal hypothesis and the predicted SM value in terms of the number of uncertainties.[1] It is clearly visible, that benchmark points with $m_A + m_H > 500$ GeV or $2\,m_{H^\pm} > 1$ TeV are below 5σ significance level, regarded as the discovery threshold, even at the highest CLIC energy, mainly due to the limited statistics of the fully-leptonic decays considered. For the charged scalar pair-production, the detection efficiency in the semi-leptonic channel is expected to be an order of magnitude higher than in

[1] The significance is the number n in the relation $|X - X_0| = n\sigma$, where X is the value measured from the experiment, X_0 is the predicted value and σ is a measurement uncertainty. For the “counting experiment”, when our result is the number of events left after the selection procedure, the expected statistical significance of signal S is given by $n = [(S+B) - B]/\sqrt{S+B} = S/\sqrt{S+B}$, where S is the expected number of signal events (eg. coming from IDM scalars production), B (background) is the number of events expected from the SM processes, $S+B$ represents the number of events measured in the experiment with $\sqrt{S+B}$ being the (Poisson) measurement uncertainty.

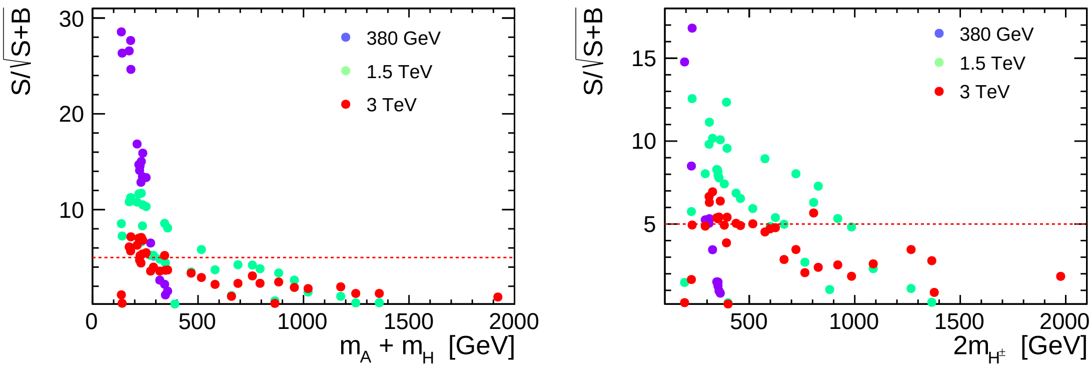

Figure 3.4: Significance of the deviations from the Standard Model predictions expected at the subsequent CLIC stages for: (left) events with two muons in the final state and (right) events with an electron and a muon in the final state, as a function of the produced scalar mass sum, for IDM benchmark models [36].

the $e^{\pm}\mu^{\mp}$ channel considered in [36]. This should result in the increase of the significance by a factor of about 3 and allow to significantly extend the IDM discovery reach of CLIC. This was the main motivation for the analysis presented in this thesis.

Chapter 4

Data processing

4.1. Signal and background samples

Sensitivity of the high energy stages of CLIC to the IDM scalar pair-production was first studied using the fast detector simulation and then verified for the selected IDM benchmark scenarios using the full detector simulation, as described in the following parts of this chapter. All event samples were generated with WHIZARD 2.7.0, assuming CLIC beam spectra and -80% polarisation of the electron beam. Background samples used in the analysis are shown in Table 4.1. Cross section values, σ, are given for both CLIC high energy stages, as well as numbers of considered Monte Carlo (MC) events for fast and full simulation analyses.

IDM benchmark scenarios considered in the full simulation study are presented in Table 4.2, together with signal cross sections, σ (corresponding to the semi-leptonic final state), at 1.5 TeV and 3 TeV, and charged dark scalar masses, $m_{H^\pm}$. Also listed are numbers of MC events used in fast and full simulation analyses. Cross sections presented in Tables 4.1 and 4.2 are given with accuracy corresponding to the statistical uncertainty of the WHIZARD integration, which is of the order of 0.1% to 1%.

4.2. Fast simulation

For fast simulation of the detector response, DELPHES [18] version 3.4.1 with CLICdet card [19] was used, as described in Sec. 2.3. This framework provides also basic high-level reconstruction of objects, such as charged leptons (electrons and muons), photons or jets (including b and tau jet tagging), and adopts particle-flow approach.

Identification of electrons, muons and photons is based on the truth, generator-level information (given in output from WHIZARD), so no misidentification is assumed [38]. Then, the candidates must pass several conditions to be stored in the final-state collections. Electrons (and positrons; $e^\pm$) and muons ($\mu^\pm$) are accepted if they propagate within the tracker acceptance and their transverse momentum is higher than threshold value of 1 GeV. Tracking detector resolution in DELPHES is taken into account by Gaussian smearing of the transverse momentum, but angular components remain unchanged. Identification efficiency cuts based on full simulation [39] are also applied, as a function of energy and pseudorapidity $\eta = -\ln[\tan(\theta/2)]$ (where θ is the polar angle), depending on particle type ($e^\pm, \mu$ or γ). Furthermore, object is considered as isolated, if in a cone of radius $\Delta R = 0.5$ surrounding it[1] sum of the transverse momenta of other particles within the cone, divided by p_T of the object, is lower than 0.12.

[1] Radius $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, where ϕ is the azimuthal angle and η is the pseudorapidity.

Table 4.1: Background event samples used in the study, together with cross sections and numbers of events used for each channel and energy stage, for fast and full simulation analysis. Background channels not considered in the full simulation study are marked with a dash.

	1.5 TeV			3 TeV		
background channel	σ [fb]	fast sim. events	full sim. events	σ [fb]	fast sim. events	full sim. events
$qq\ell\nu$	7060	1,000,000	1,010,898	8830	1,000,000	968,976
$qqqq$	1970	1,000,000	118,800	917	1,000,000	210,552
$qq\ell\ell$	2740	1,000,000	979,073	3220	1,000,000	938,592
$\ell\ell$	1400	1,000,000	485,199	1670	1,000,000	400,632
$qq\nu\nu$	1520	1,000,000	-	2510	1,000,000	-
qq	2870	1,000,000	-	1230	1,000,000	-
$qqqq\ell\nu$	177	100,000	-	131	100,000	-
$qq\ell\nu\ell\nu$	56.9	100,000	-	63.3	10,000	-
$qq\ell\nu\nu\nu$	40.2	100,000	-	70.8	10,000	-

Table 4.2: Signal event samples used in the study, together with cross sections and numbers of events used in fast and full simulation, and charged scalar masses, for benchmark points selected for the full simulation analysis. Scenarios with off-shell $W^{\pm}$ production are marked with a star.

	1.5 TeV			3 TeV			
Benchmark point	σ [fb]	fast sim. events	full sim. events	σ [fb]	fast sim. events	full sim. events	$m_{H^{\pm}}$ [GeV]
BP21	8.19	100,000	49,500	4.27	100,000	48,024	299.536
BP23	12.6	100,000	49,500	5.68	100,000	45,552	190.822
HP3*	0.634	100,000	49,500	1.73	100,000	45,600	633.48
HP17*	2.45	100,000	49,500	1.42	100,000	46,056	440.624
HP20*	1.33	100,000	49,500	1.54	100,000	43,200	543.794

Jets are reconstructed using the Valencia Linear Collider (VLC) algorithm [40], which is the most accurate for the high-energy lepton colliders and, as a part of work on CLICdet card, was added permanently to the Delphes framework. Jets are clustered with FastJet package [41] in two modes – inclusive and exclusive. In the inclusive mode, algorithm merges objects given only the radius, and clusters as many jets as it is able to. On the other hand, exclusive mode forces the algorithm to cluster fixed number of jets with the given radius. By default, in CLICdet card jets are clustered with the VLC parameters $\gamma = \beta = 1$ in both modes, with several R values ranging from 0.5 to 1.5 combined also with (in the exclusive mode) number of jets from 2 to 6. Concerning jet reconstruction, one also has to take into account possible effects of the beam-induced background. The influence of soft $\gamma\gamma \to$ hadrons events (so-called *overlay* background) becomes important at high energy stages of CLIC and is not modelled in Delphes. To take it into account, additional jet energy smearing is applied, depending on the jet angle. This approach is very simplified, as the background is also expected to affect the jet clustering, while the energy smearing is applied on the already clustered jets. However, more sophisticated description of the overlay background was not possible in the existing CLICdet detector model for Delphes.

4.3. Full simulation

After the fast simulation analysis covering all benchmark scenarios considered in [33, 36], full simulation study was performed for the selected scenarios, as listed in Table 4.2. Event samples were generated again with WHIZARD 2.7.0, using PYTHIA6 for hadronization. Then, detector response for CLICdet model *CLIC_o3_v14* was simulated with GEANT4, and DD4hep toolkit [42] served for the description of the detector. The PandoraPFA algorithm [12] is the key element of the event reconstruction. Based on combined information from calorimeters, tracking detectors and muon systems, the collections of the Pandora Particle Flow Objects (PandoraPFOs or just PFOs) are created, which contain four-momenta of the reconstructed particles.

Beam-induced backgrounds, expected at CLIC due to high beam intensity and high bunch repetition rate, were included in the full simulation. These processes can result in additional tracks, hits or energy deposits in the detector, that may be misidentified as objects resulting from the primary hard interaction, used in the particle flow reconstruction and saved in the PFO collections. For this reason, an algorithm with additional background-rejection criteria is run on the reconstructed PFOs, with cuts depending on the expected running conditions. Three standard sets of cuts are used in CLICdp studies (loose, medium and tight selection), corresponding to the three considered energy stages – the higher energy, the tighter selection has to be used.

Particle Flow Objects are then identified as particular particles or used to reconstruct composite objects. The first step is the identification of isolated leptons, which is performed with IsolatedLeptonFinder Marlin processor [43]. Minimum track energy to classify object as an isolated lepton was 5 GeV. Next, the so-called polynomial isolation criterion was applied. It is based on a relation between the considered track energy, E_{track}, and the energy of all other particles in a cone surrounding the track, E_{cone}, for a cone apex angle θ corresponding to $\cos(\theta/2) = 0.99$. The track was considered as an isolated lepton, if

$$E_{cone}^2 < A \cdot E_{track}^2 + B \cdot E_{track} + C \ ,$$

where A, B and C are parameters of the algorithm. The values of $A = 4.43$, $B = -22.2$ and $C = 333.3$ were chosen for the presented study to maximize, as much as possible, the efficiency of isolated leptons reconstruction. Performance of the algorithm was verified by matching the objects, that passed the algorithm selection, with true, generator-level leptons, within the detector acceptance. The parameters' values choice was optimized by matching PFOs with true leptons and comparing their E_{cone}-E_{track} dependence with the same relation for misidentified isolated leptons. For lepton flavour identification, types given by PandoraPFA were used. At last, leptons were also dressed with close-by photons and electrons (merged into higher energy lepton).

Pandora PFOs not classified as isolated leptons were subsequently used for jets clustering. It was done again with VLC algorithm, using Marlin FastJet processor, requiring two jets in the exclusive mode with $\gamma = \beta = 1$ and several R parameter values to allow further tests.

All kinematic variables of the reconstructed objects, including identified leptons and results of jet clustering, as well as all other accessible, useful information on the events, were later saved in a ROOT [22] tree using Marlin processor, based on the processor developed previously for WW production analysis [44].

Chapter 5

Fast simulation analysis

The main goal of the presented study is to verify sensitivity of high energy stages of CLIC to IDM charged scalar pair-production in the semi-leptonic channel, when one expects isolated electron or muon in the final state (tau reconstruction is not considered) and a pair of jets. This channel is expected to give an order of magnitude higher statistics than the $e^{\pm}\mu^{\mp}$ channel considered in [36].

5.1. Event preselection

The first step in the analysis was to make the cut-based event preselection. A presence of exactly one isolated lepton (μ or e) and two jets were required in each event. On the generator-level, all three lepton flavours were generated in the final state. It means, that maximum possible preselection efficiency is 78%, as only leptonic tau decays give the required signature and hadronic tau decays are rejected at this point. Jets were reconstructed with VLC algorithm in the exclusive mode, with $\gamma = \beta = 1$, $R = 1.0$ at 1.5 TeV and $R = 1.2$ at 3 TeV. For both energy stages, events containing an isolated photon with energy $E_\gamma > 10\,\text{GeV}$ were rejected. Additional energy in the event, that is the energy of all particle flow objects not contributing to the expected signature, was required to be below 20 GeV.

Subsequently, the cut-based selection was applied. Distributions of the kinematic variables describing the considered final state and used for the preselection cuts are presented in Figures 5.1 and 5.2, for 1.5 TeV and 3 TeV respectively. The plots show histograms for the two selected IDM scenarios (BP21 and HP17) and all the considered SM background channels summed. Selection cuts were chosen to obtain signal selection efficiency of at least 30% with the best resulting signal to background ratio. For the 1.5 TeV energy stage following conditions were imposed:

$$\begin{array}{llll} M_{jj} > 3\,\text{GeV}, & E_{j_1} + E_{j_2} < 750\,\text{GeV}, & E_\ell < 600\,\text{GeV}, & p_T^\ell < 550\,\text{GeV}, \\ M_{miss} > 400\,\text{GeV}, & 0.2 < \theta_W < \pi - 0.2, & 0.25 < \theta_\ell < \pi - 0.25. & \end{array} \tag{5.1}$$

Here M_{jj} is the invariant mass of a dijet system (so the invariant mass of hadronically decaying W boson), E_{j_1} and E_{j_2} are the energies of each of the jets, E_ℓ and p_T^ℓ are respectively the energy and the transverse momentum of an isolated lepton, θ_W is the polar angle of hadronically decaying W (a dijet system) and θ_ℓ is the polar angle of an isolated lepton. The missing mass, M_{miss}, was calculated by taking the four-momentum of the initial e^+e^- state ($\vec{p} = \vec{0}$, $E = \sqrt{s}$) and subtracting the four-momenta of the isolated lepton and of the two jets; the invariant mass corresponding to the remaining four-momentum is the missing mass.

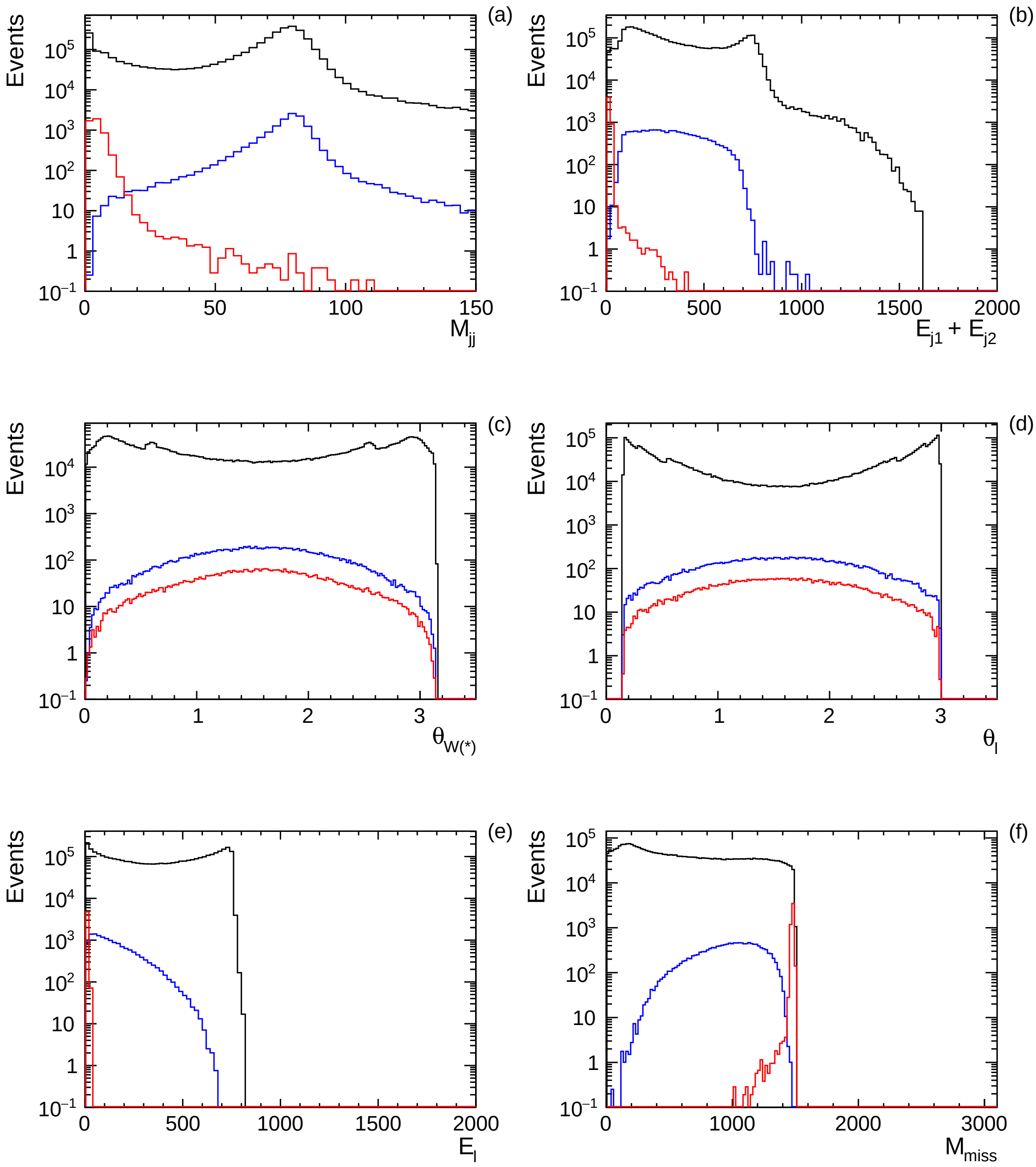

Figure 5.1: Histograms of the masses of the dijet system (a), sums of the energy of two jets (b), polar angle of a dijet system (c), polar angle of a lepton (d), energy of a lepton (e) and missing mass (f) at $\sqrt{s} = 1.5$ TeV. The red line denotes HP17, and the blue one BP21. Black histogram is the sum of all background channels. Normalization is the number of events expected in the real experiment.

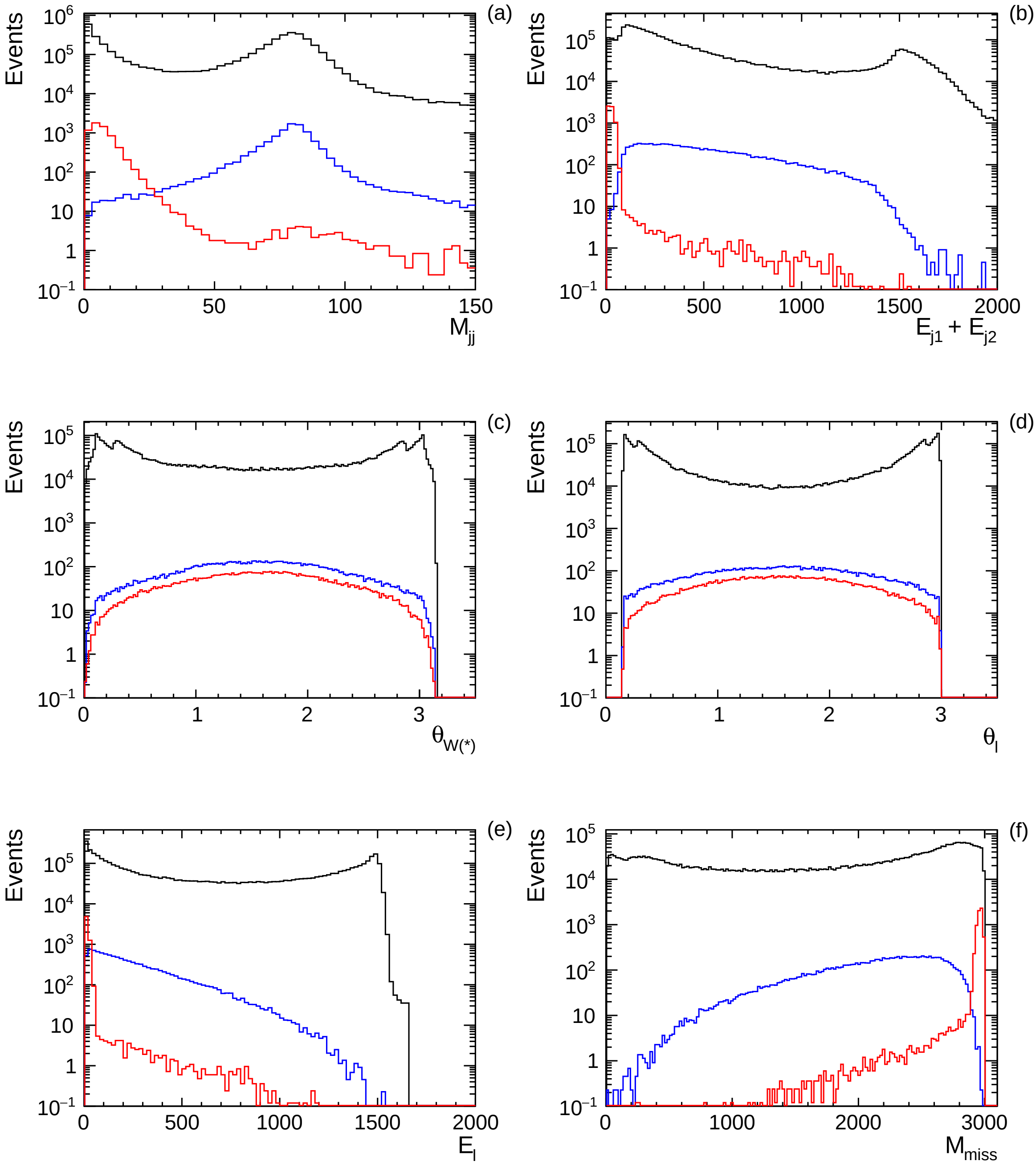

Figure 5.2: Histograms of the masses of a dijet system (a), sums of energy of two jets (b), polar angle of a dijet system (c), polar angle of a lepton (d), energy of a lepton (e) and missing mass (f) at $\sqrt{s} = 3$ TeV. The red line denotes HP17, and the blue one BP21. Black histogram is the sum of all background channels. Histograms are normalized to the number of events expected in the real experiment.

Table 5.1: Results of the preselection for the two exemplary benchmark points and four dominating background channels. The table shows all expected events (in the semi-leptonic channel) and expected events after preselection cuts, as well as total preselection efficiency.

$\sqrt{s} = 1.5\,\text{TeV}$			
channel	all exp. ev.	exp. ev. after preselec.	efficiency
H^+H^- (BP21)	16,400	9,400	57%
H^+H^- (HP17)	4,890	1,580	32%
$qq\ell\ell$	5,480,000	161,000	2.9%
$qq\ell\nu$	14,100,000	832,000	5.9%
$qq\ell\nu\ell\nu$	114,000	14,800	13%
$qq\ell\nu\nu\nu$	80,300	30,800	38%
total background	35,700,000	1,080,000	3%
signal/backg. (BP21)	0.00046	0.0087	
signal/backg. (HP17)	0.00014	0.0015	
$\sqrt{s} = 3\,\text{TeV}$			
channel	all exp. ev.	exp. ev. after preselec.	efficiency
H^+H^- (BP21)	17,100	7,020	41%
H^+H^- (HP17)	5,690	2,200	39%
$qq\ell\ell$	12,900,000	78,400	0.61%
$qq\ell\nu$	35,300,000	400,000	1.1%
$qq\ell\nu\ell\nu$	317,900	30,700	9.7%
$qq\ell\nu\nu\nu$	361,000	63,600	18%
tot. backg.	74,900,000	626,000	0.84%
signal/backg. (BP21)	0.00023	0.011	
signal/backg. (HP17)	0.000076	0.0035	

For the analysis at 3 TeV energy stage, accepted were events with

$$\begin{array}{c} M_{jj} > 3\,\text{GeV}, \quad E_{j_1} + E_{j_2} < 1.3\,\text{TeV}, \qquad E_\ell < 1\,\text{TeV}, \qquad p_T^\ell < 800\,\text{GeV} \\ 0.3 < \theta_W < \pi - 0.3, \quad 0.5 < \theta_\ell < \pi - 0.5. \end{array} \tag{5.2}$$

Detailed information about the number of events before and after the preselection, efficiencies, and the signal to background ratios, for the two considered signal scenarios and the dominating background channels, are presented in Table 5.1, for 1.5 TeV and 3 TeV energy stages. The table presents numbers of expected events for the total integrated luminosities of $2\,\text{ab}^{-1}$ at 1.5 TeV and $4\,\text{ab}^{-1}$ at 3 TeV assumed for running with -80% electron beam polarisation [45].

One can see, that it was possible to strongly reduce amount of the background and give raise of about one to two orders of magnitude to the signal to background ratio. Worth noting is also, that the same type of cuts is much more affecting the samples at the 3 TeV, which results in lower signal efficiency, but at the same time, even higher background reduction. This is allowed by longer tails in the background distributions, as well as more forward (in terms of polar angle) flight direction of the leptons and hadronic $W^\pm$, at the higher energy stage. What is more, larger cuts on the angular variables at the 1.5 TeV led to insufficient signal selection efficiency. At last, scenarios with off-shell $W^\pm$ production tend to have much lower values of energy and momentum of leptons and jets (because of smaller $m_{H^\pm} - m_H$ difference), which makes them more difficult to separate from the background. It also results in lower selection efficiency for these scenarios already at the stage of requiring the proper signature.

Preselection was also optimised taking into account the results from multivariate analysis (the next step, Sec. 5.2). It is not the "rule of thumb", that the MVA works best given the highest signal to background ratio, if the price to pay is to have much lower statistics in its input data.

5.2. Multivariate analysis

The preselection was clearly not enough to extract the signal from the data. Afterwards, the multivariate analysis technique was used. In this study, the BDT algorithm, provided by TMVA Toolkit [21], was trained separately on the two datasets – one including all signal scenarios with an on-shell $W^{\pm}$, and another with an off-shell $W^{\pm}$ boson production. It can be seen in Figures 5.1 and 5.2, that this choice is justified by the huge difference between distribution shapes of these two classes of scenarios. One thousand decision trees were used for the training, with eleven input variables: M_{jj}, $E_{jj} \equiv E_{j_1} + E_{j_2}$, E_{ℓ}, p_T^{ℓ}, θ_W, θ_{ℓ}, M_{miss}, $\Delta\theta^*_{Wj}$, $\Delta\phi^*_{Wj}$, E_{flow} and E_T^{miss}. Variables $\Delta\theta^*_{Wj}$ and $\Delta\phi^*_{Wj}$ are respectively polar and azimuthal angles between the jet and the flight direction of (hadronically decaying) $W^{\pm}$, calculated in its centre-of-mass frame, E_{flow} is the total energy of additional objects in the detector, not contributing to the required signature, and E_T^{miss} is the missing transverse energy (MET). MET is calculated as

$$\overrightarrow{E}_T^{miss} = -\sum_i \overrightarrow{p}_T(i), \tag{5.3}$$

where sum goes over all reconstructed particle-flow objects in the detector.

BDT response distribtions, for the two exemplary scenarios, BP21 and HP17, are presented in Figure 5.3. BDT response is expected to take values from -1 to +1: the closer is the value to +1, the more likely the considered event is a signal event, and the closer is the response to -1, the higher is the probability of background hypothesis. One is then able to look for a cut on the BDT variable, that gives the highest possible signal significance. This procedure was also done using the TMVA Toolkit.

Presented in Figure 5.3 are the BDT output distributions both for the training and the validation (test) background samples. Results from the training match those from the test, which means, that the algorithm was not *overtrained*. Overtraining occurs, when a program learns the specific dataset too well and becomes sensitive to statistical fluctuations, which is visibly not the case here. For the signal, training sample included all the IDM scenarios (on- or off-shell), while only the selected scenario was considered for the test sample. That is why similar comparison for the signal sample is not possible here.

Total selection efficiency after all cuts, for all considered scenarios, was required to be at least 10%. Cross sections, σ, efficiencies, ϵ_{pre} and $\epsilon_{BDT>X}$, of the preselection and the multivariate-based selection (respectively), as well as the expected number of events, $N_{BDT>X}$, after all the cuts are presented in Table 5.2, for the 1.5 TeV and the 3 TeV CLIC energy stages, for the two selected scenarios, BP21 and HP17. Optimal cut on the BDT response, $BDT > X$, was found by TMVA separately for each scenario. The table shows also efficiencies and numbers of expected events for the background, that also depend on the considered scenario (due to BDT response cut optimisation).

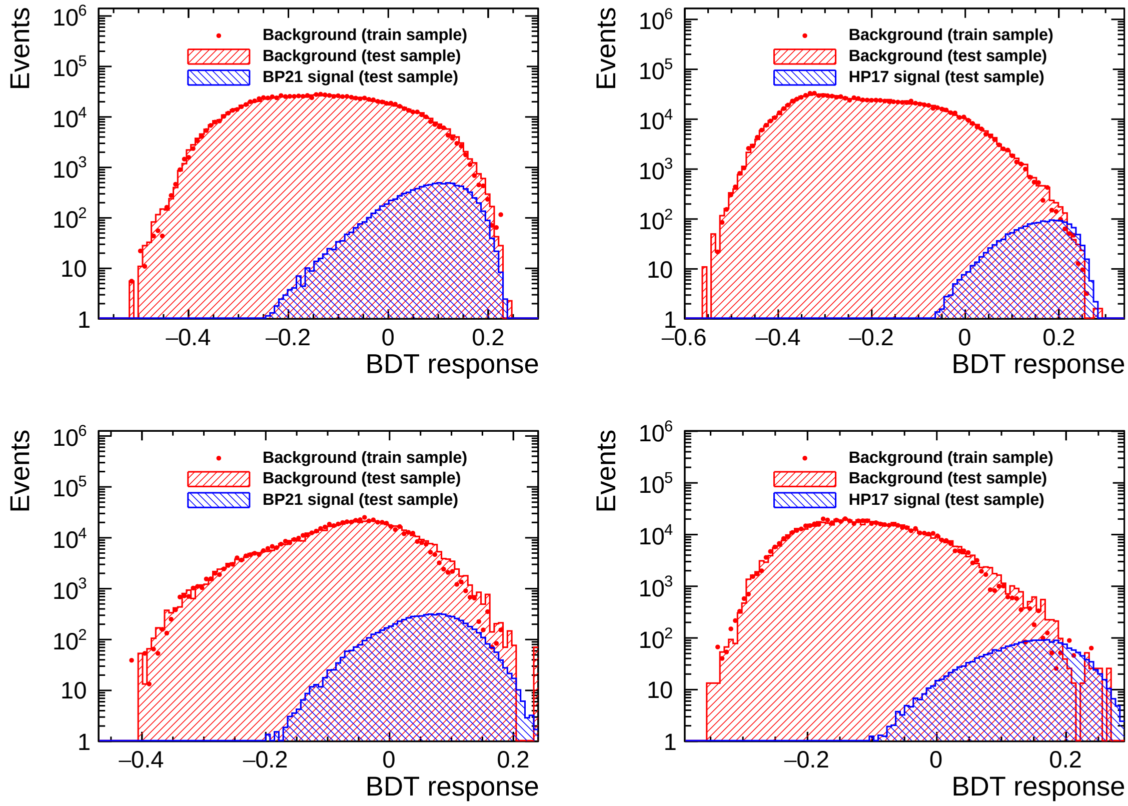

Figure 5.3: BDT response distributions for the BP21 (left) and the HP17 (right) scenarios at $\sqrt{s} = 1.5\,\text{TeV}$ (upper plots) and at $\sqrt{s} = 3\,\text{TeV}$ (bottom plots). Benchmark scenarios and SM background expectations are normalised to the integrated luminosity of $2\,\text{ab}^{-1}$ and $4\,\text{ab}^{-1}$, for 1.5 TeV and 3 TeV, respectively.

5.3. Results

Statistical significance obtained after the cut on the BDT response variable, for given benchmark point, is considered here as the final result. Results for all considered benchmark scenarios and the two CLIC high energy stages, are presented as a function of $2\,m_{H^\pm}$ in Figure 5.4. Results of the leptonic channel study [36] are also included for comparison. It can be seen, that all the scenarios could be potentially observed at 3 TeV energy stage, besides one (HP10). At 1.5 TeV four scenarios are below the discovery threshold, and the other four are beyond the kinematic reach of the running stage ($m_{H^\pm} > 750\,\text{GeV}$). All the scenarios with on-shell $W^\pm$ production could be detected; resulting expected significance is high, it reaches even over 60σ. The discovery reach of CLIC for high mass charged IDM scalars is significantly extended for the semi-leptonic channel as compared to the leptonic channel sensitivity.

Full set of the results is presented in Table A.2 in the Appendix. Significances expected for all scenarios are listed together with signal cross sections and charged dark scalar masses.

Table 5.2: Efficiencies of preselection and cuts on BDT variable, together with expected number of events after respective cut, for $\sqrt{s} = 1.5$ TeV and $\sqrt{s} = 3$ TeV, two selected benchmark points and the background. Best BDT selection cut was found by TMVA separately for each scenario.

$\sqrt{s} = 1.5$ TeV			on-shell W		off-shell $W^\star$	
Sample	σ [fb]	ϵ_{pre}	$\epsilon_{BDT>0.08}$	$N_{BDT>0.08}$	$\epsilon_{BDT>0.17}$	$N_{BDT>0.17}$
BP21	8.19	57%	55%	5190	-	-
HP17	2.45	32%	-	-	45%	714
background	17800	3%	5%	53500	0.13%	1310
$\sqrt{s} = 3$ TeV						
Sample	σ [fb]	ϵ_{pre}	$\epsilon_{BDT>0.07}$	$N_{BDT>0.07}$	$\epsilon_{BDT>0.17}$	$N_{BDT>0.17}$
BP21	4.27	41%	45%	3120	-	-
HP17	1.42	39%	-	-	35%	767
background	18700	0.84%	5.3%	33300	0.14%	846

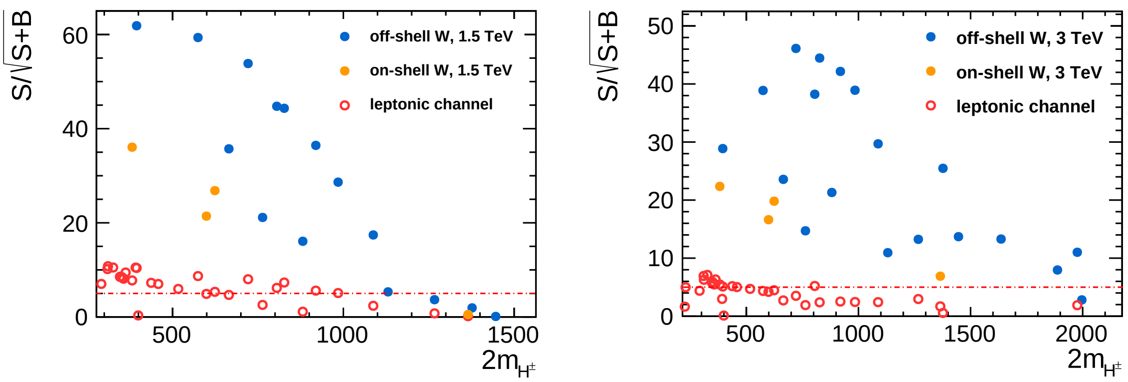

Figure 5.4: Resulting statistical significance for all scenarios as a function of $2m_{H^\pm}$ at 1.5 TeV (left) and 3 TeV (right), both for the scenarios with on-shell $W^\pm$ boson and off-shell (orange) $W^\pm$ boson (blue). Open red circles show the expected results for leptonic channel [36], and the red dotted line denotes 5σ threshold.

Chapter 6

Full simulation analysis

6.1. Event preselection

After running the full simulation of detector response for selected benchmark scenarios, the further analysis was carried out in the same way, as for the simulation with Delphes. In the beginning, a regular, cut-based event preselection was applied. It was meant to be as similar as possible to the one applied for the fast simulation. This would assure, that the two studies are more comparable, and allow to treat this analysis also as a form of validation of the previous one. The main difference between the two approaches is a presence of $\gamma\gamma \rightarrow$ hadrons background in the full simulation (*overlay* events), which is not included in Delphes, so the biggest challenge was to minimize its influence on the results.

Each event was required to contain exactly one isolated lepton and a pair of jets, reconstructed in the exclusive mode, with the VLC algorithm. VLC parameters were fixed to $\gamma = \beta = 1$, $R = 0.9$ for the analysis at the 1.5 TeV energy stage and $R = 1.2$ at 3 TeV. No isolated photons with energy $E_\gamma > 10$ GeV were allowed; a photon with transverse momentum p_T^γ was considered as isolated, if $p_T^{cone}/p_T^\gamma \leq 0.2$, where p_T^{cone} is the transverse momentum of all particles in a cone surrounding photon direction, defined by squared radius $(\Delta R)^2 = (\Delta\theta)^2 + \sin^2(\theta) \cdot (\Delta\phi)^2 < 0.1$. Events with total of transverse momentum of additional (with respect to the required event signature) particle-flow objects, p_T^{flow}, greater than 20 GeV were also removed from the analysis. This requirement, $p_T^{flow} < 20$ GeV, replaces the condition of $E_{flow} < 20$ GeV used in the fast simulation analysis. This is because E_{flow} turned out to be highly influenced by the overlay, leading to a very long tail in its distribution, which makes the cut much less efficient. The p_T^{flow} distribution is much less affected by the overlay background as most of the additional energy deposits are of rather small transverse momentum.

Selection based on the kinematic variables, describing the system of pair of jets, lepton and a missing energy, was further applied. The distributions of these variables are presented in Figures 6.1 and 6.2, for 1.5 TeV and 3 TeV CLIC running stages respectively, for the selected signal scenarios and the background. Corresponding distributions from the fast simulation study are also included for comparison. The differences in the normalization reflect the differences in the event selection criteria and the influence of the beam-induced backgrounds (selection of isolated photons and additional particle-flow objects have different efficiencies). Overlay background is also responsible for the visible smearing of the histograms, as well as for the flattening of the angular distributions of dijet systems for the scenario with off-shell $W^\pm$.

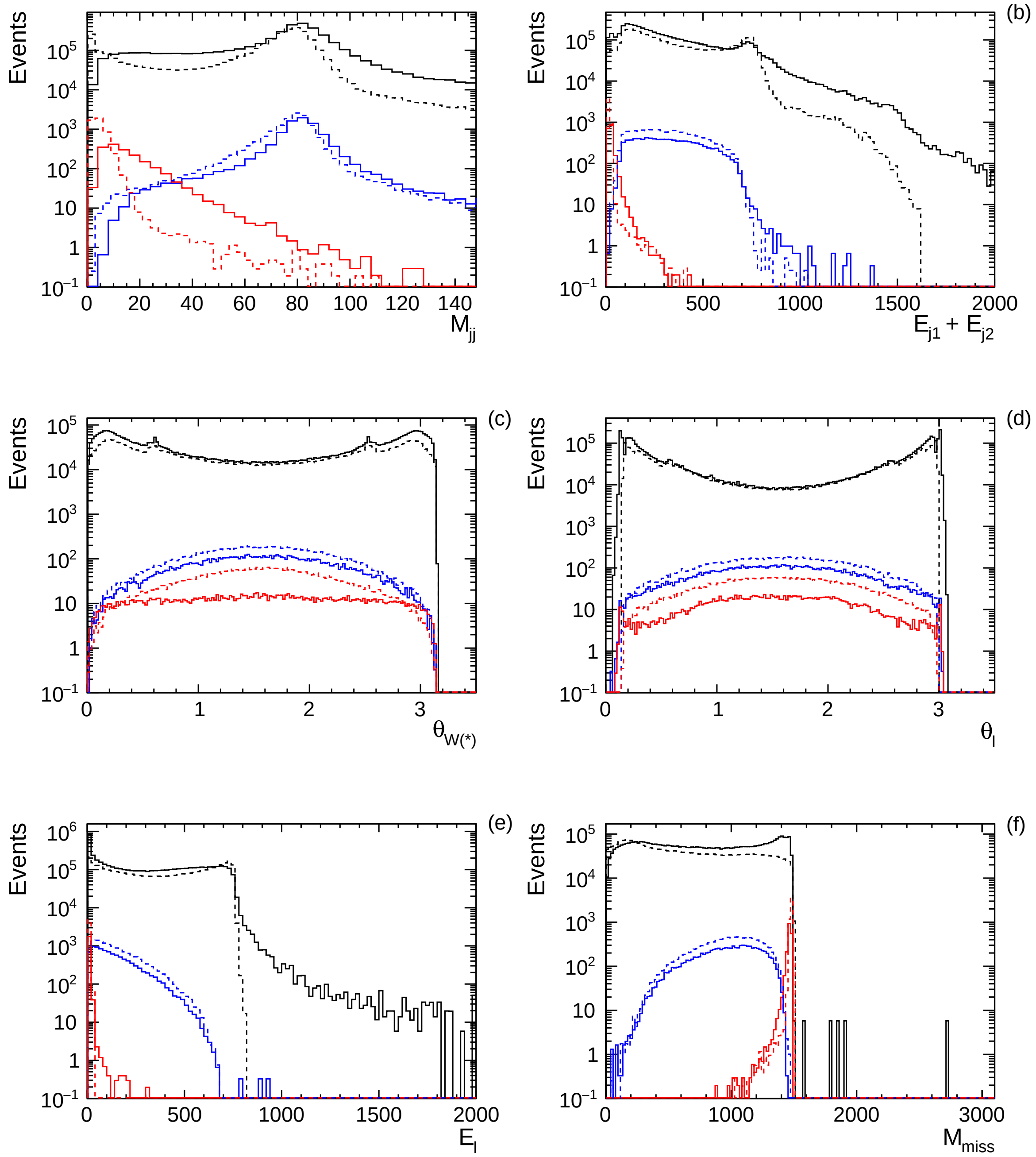

Figure 6.1: Histograms of the masses of the dijet system (a), sums of the energy of two jets (b), polar angle of a dijet system (c), polar angle of a lepton (d), energy of a lepton (e) and missing mass (f) at $\sqrt{s} = 1.5$ TeV. The red line denotes HP17, and the blue one BP21. Black histogram is the sum of all background channels. Normalization is the number of events expected in the real experiment. Solid line shows the full simulation histograms, and the dotted line the corresponding fast simulation ones.

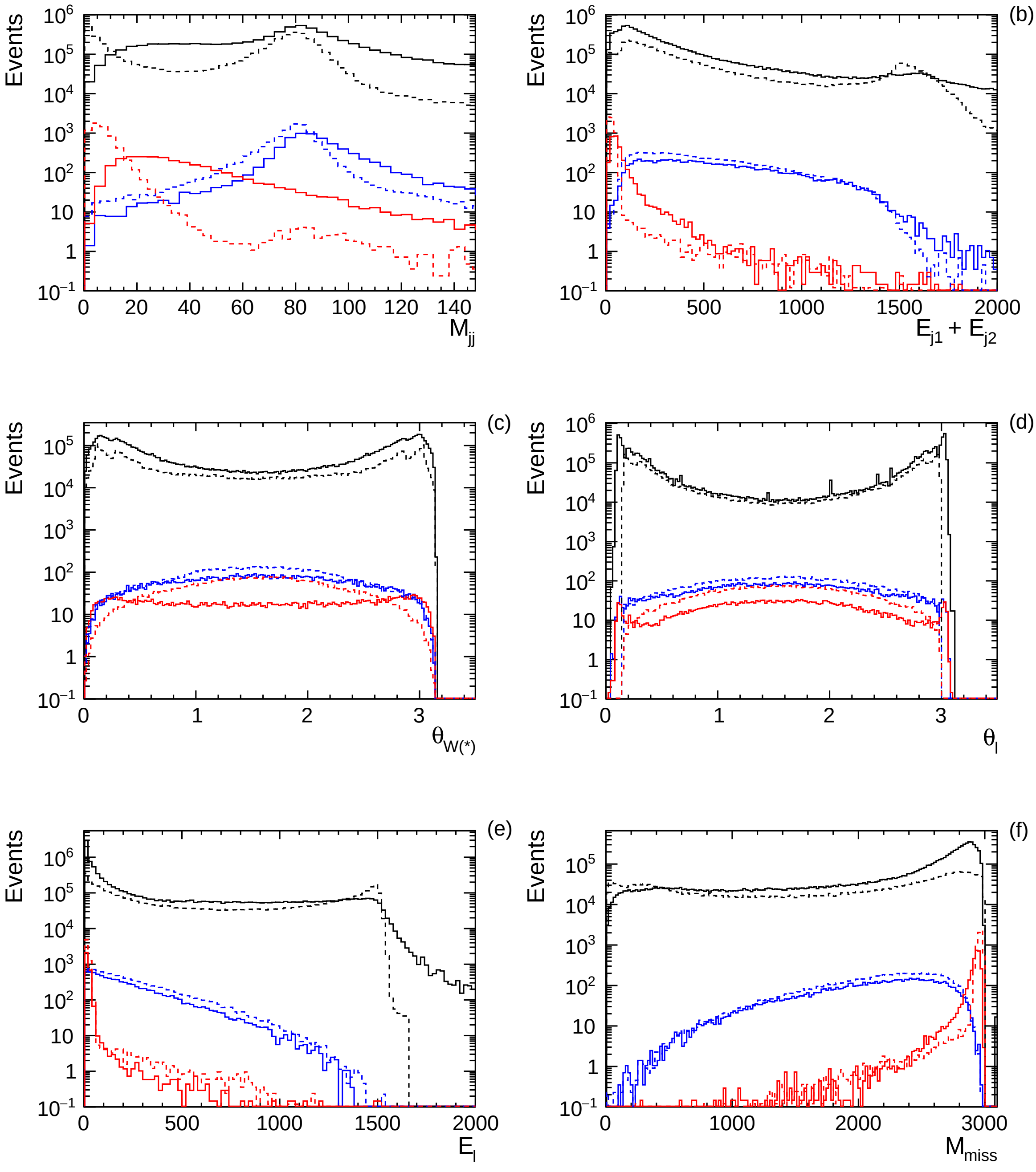

Figure 6.2: Histograms of the masses of the dijet system (a), sums of the energy of two jets (b), polar angle of a dijet system (c), polar angle of a lepton (d), energy of a lepton (e) and missing mass (f) at $\sqrt{s} = 3$ TeV. The red line denotes HP17, and the blue one BP21. Black histogram is the sum of all background channels. Normalization is the number of events expected in the real experiment. Solid line shows the full simulation histograms, and the dotted line the corresponding fast simulation ones.

Table 6.1: Results of the preselection for the two exemplary benchmark points and four background channels considered in full simulation study. The table shows numbers of all expected events (in the semi-leptonic channel) and expected events after preselection cuts, as well as total preselection efficiency.

$\sqrt{s} = 1.5\,\text{TeV}$			
channel	all exp. ev.	exp. ev. after preselec.	eff.
H^+H^- (BP21)	16,200	9100	56%
H^+H^- (HP17)	4860	1590	33%
$qq\ell\ell$	5,430,000	232,000	4.3%
$qq\ell\nu$	14,000,000	1,000,000	7.1%
$\ell\ell$	2,810,000	140,000	5%
$qqqq$	3,890,000	12,000	0.31%
tot. backg.	26,100,000	1,386,063	5.3%
signal/backg. (BP21)	0.00062	0.0084	
signal/backg. (HP17)	0.00019	0.0015	
$\sqrt{s} = 3\,\text{TeV}$			
channel	all exp. ev.	exp. ev. after preselec.	eff.
H^+H^- (BP21)	16,900	6460	38%
H^+H^- (HP17)	6710	2020	30%
$qq\ell\ell$	12,700,000	203,000	1.6%
$qq\ell\nu$	34,700,000	707,000	2%
$\ell\ell$	6,660,000	83,100	1.3%
$qqqq$	3,610,000	14,900	0.41%
tot. backg.	57,700,000	1,010,000	1.8%
signal/backg. (BP21)	0.00029	0.0064	
signal/backg. (HP17)	0.00012	0.002	

The following preselection cuts on these variables were set for the 1.5 TeV energy stage:

$$\begin{gathered} M_{jj} > 3\,\text{GeV},\ E_{j_1} + E_{j_2} < 750\,\text{GeV},\ E_\ell < 600\,\text{GeV},\ 2\,\text{GeV} < p_T^\ell < 550\,\text{GeV}, \\ M_{miss} > 400\,\text{GeV}, \quad 0.2 < \theta_W < \pi - 0.2, \quad 0.25 < \theta_\ell < \pi - 0.25, \end{gathered} \tag{6.1}$$

and for the 3 TeV stage:

$$\begin{gathered} M_{jj} > 2\,\text{GeV},\ E_{j_1} + E_{j_2} < 1.3\,\text{TeV},\ E_\ell < 1\,\text{TeV},\ p_T^\ell < 800\,\text{GeV} \\ 0.3 < \theta_W < \pi - 0.3, \quad 0.5 < \theta_\ell < \pi - 0.5. \end{gathered} \tag{6.2}$$

Presented in Table 6.1 are preselection results for the considered full simulation samples, analogous to those in Table 5.1 for the fast simulation. Numbers of events expected before the preselection are different from the numbers in Table 5.1, because the assumed cross sections (that served to calculate them) computed by WHIZARD were slightly different for the produced samples for fast and full simulation. These discrepancies can result from small differences in generator level cuts but are also compatible with the expected statistical uncertainty of the WHIZARD cross section integration.

It can be seen, that preselection efficiencies for the signal are in most cases lower than for the fast detector simulation, while for the background they are higher. It is most likely due to the larger smearing in the distributions of kinematic variables (cf. Figures 6.1 and 6.2), which can result from the overlay backgrounds but also from more detailed particle tracking and detector response simulation. Maintaining high selection efficiency in the presence of overlay background is more difficult for the scenarios with off-shell $W^{\pm}$ production. Energy and momentum of isolated leptons decrease with the $W^{\pm}$ virtuality and their reconstruction can be significantly affected by the beam-induced background. For this reason, efficiency, compared to the fast simulation, drops significantly already on the event classification stage (selecting particular final state). Still, similar to the preselection stage in the fast simulation analysis, it was possible to improve the signal-to-background ratio significantly.

6.2. Multivariate analysis

After the cut-based preselection, the BDT algorithm provided by TMVA Toolkit was utilized, following the same procedure as used in the fast simulation analysis. One thousand decision trees were trained, applying Gaussianization and variable decorrelation on the input data. Training was performed separately on the two on-shell $W^{\pm}$ scenarios, and on the three samples with off-shell $W^{\pm}$ production. The following input variables were used for training: M_{jj}, E_{jj}, E_{j_1}, E_{j_2}, $p_T^{j_1}$, $p_T^{j_2}$, E_ℓ, p_T^ℓ, θ_W, θ_{j_1}, θ_{j_2}, θ_ℓ, M_{miss}, $\Delta\theta^*_{Wj}$, $\Delta\phi^*_{Wj}$, E_{flow}, E_T^{miss}. This time, the kinematic variables of particular jets were also used – energy, $E_{j_{1,2}}$, transverse momentum, $p_T^{j_{1,2}}$ and polar angle, $\theta_{j_{1,2}}$.

BDT response distributions, for the selected scenarios and the two energy stages considered, are presented in Figure 6.3. Slight overtraining might be noticed when comparing to the fast simulation results – as fewer signal samples were used for the training – but this effect is not significant.

Results of the selection – efficiencies of the preselection and of the BDT cut, as well as expected numbers of events after the selection, for the two exemplary signal scenarios and for the SM background, are presented in Table 6.2, together with the corresponding cross section values. The optimal value of the BDT response cut, maximizing the expected statistical significance, was found by TMVA. As for the fast simulation study, total selection efficiency was required to be at least 10%.

Table 6.2: Efficiencies of preselection and cuts on BDT variable, together with expected number of events after respective cut, for $\sqrt{s} = 1.5$ TeV and $\sqrt{s} = 3$ TeV, two selected benchmark points and the background. Best BDT selection cut was found by TMVA for each scenario.

$\sqrt{s} = 1.5$ TeV						
Sample	σ [fb]	ϵ_{pre}	$\epsilon_{BDT>0.06}$	$N_{BDT>0.06}$	$\epsilon_{BDT>0.11}$	$N_{BDT>0.11}$
BP21	8.19	56%	54%	4940	-	-
HP17	2.4	33%	-	-	32%	514
background	13100	5.3%	3.1%	43100	0.17%	2400
$\sqrt{s} = 3$ TeV						
Sample	σ [fb]	ϵ_{pre}	$\epsilon_{BDT>0.12}$	$N_{BDT>0.12}$	$\epsilon_{BDT>0.09}$	$N_{BDT>0.09}$
BP21	4.27	38%	33%	2130	-	-
HP17	1.42	30%	-	-	44%	883
background	14400	0.41%	0.92%	9280	1.6%	16500

6.3. Results

The TMVA framework was used to calculate the maximum expected deviations from the SM predictions observed in the experiment for each considered IDM scenario. These results

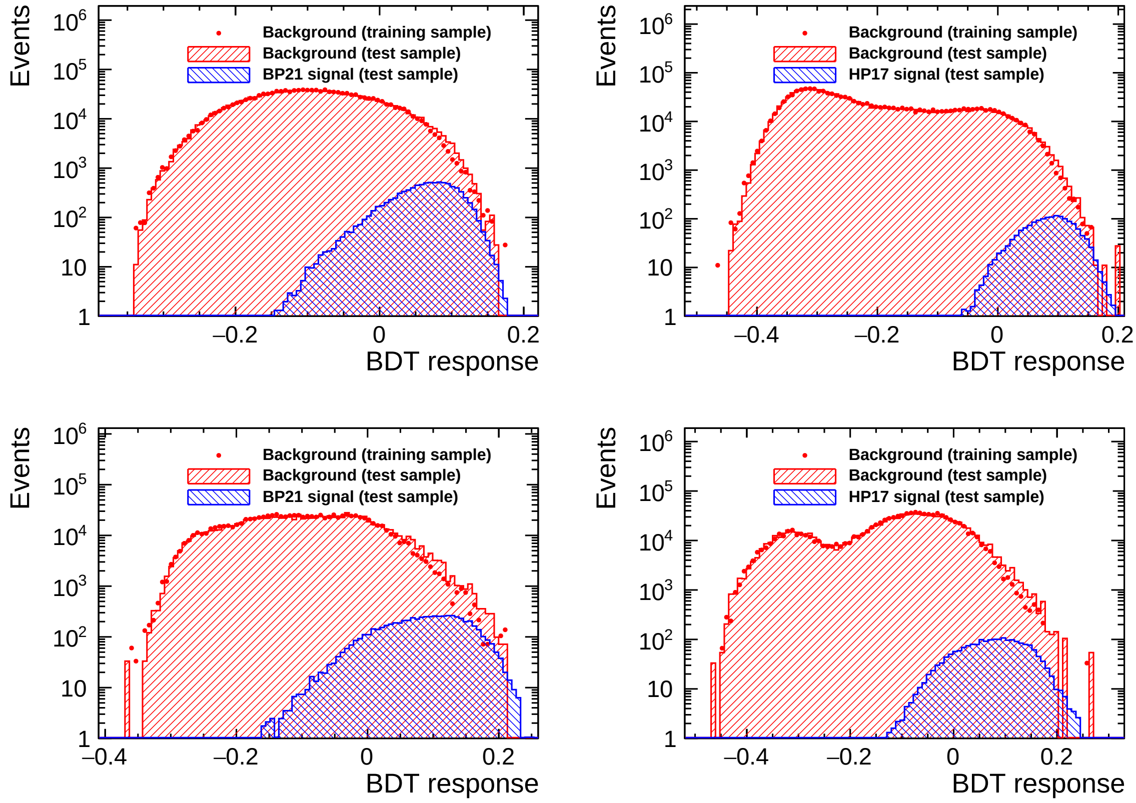

Figure 6.3: BDT response distributions for the BP21 (left) and the HP17 (right) scenarios at $\sqrt{s} = 1.5\,\text{TeV}$ (upper plots) and at $\sqrt{s} = 3\,\text{TeV}$ (bottom plots). Benchmark scenarios and SM background expectations are normalised to the integrated luminosity of $2\,\text{ab}^{-1}$ and $4\,\text{ab}^{-1}$, for 1.5 TeV and 3 TeV, respectively.

are presented in Figure 6.4 and the values for two exemplary scenarios in Table 6.3. The corresponding results from the fast simulation analysis are also included for comparison in Figure 6.4.

The obtained results are visibly worse compared to the fast detector simulation study, but still all the considered points (besides one, HP3, at 1.5 TeV) could be potentially discovered at CLIC. It can be seen, that significance drops much more for the scenarios with off-shell $W^{\pm}$ production, while for those with on-shell $W^{\pm}$ (the two low-mass points) it remains almost unchanged. However, one has to take into account that not all background channels were included in the full simulation analysis. Significances resulting from the fast simulation, with the same background contribution as in the full simulation, are also included for comparison in Figure 6.4.

6.4. Validation of fast simulation

One of the goals of the full simulation study was to validate, as much as possible, the fast simulation analysis. Presented results indicate that the simulation based on the DELPHES framework can only be considered as reliable for scenarios with real $W^{\pm}$ production. For scenarios with virtual $W^{\pm}$ production, results from fast simulation are too optimistic – much lower statistical significance is obtained with full simulation. The difference between results

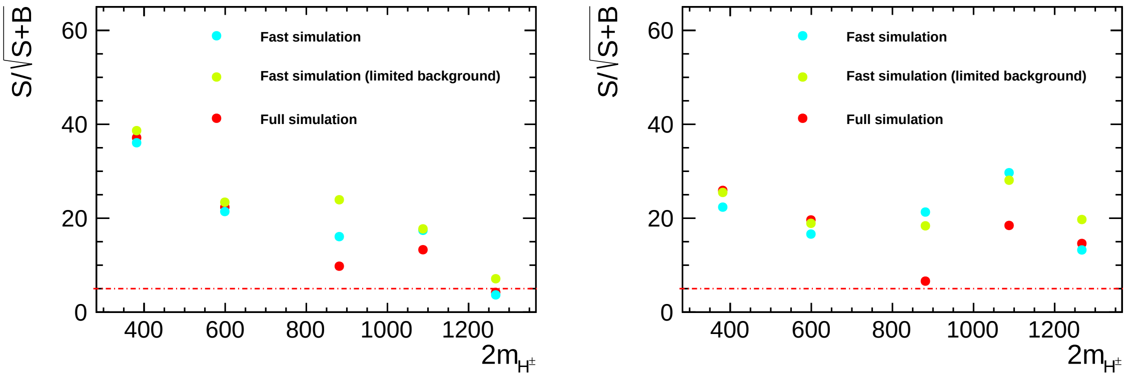

Figure 6.4: Statistical significance expected from the full simulation study (red points), compared with the fast simulation results (blue points). Results from the fast simulation analysis with the background samples limited to those used in the full simulation analysis are shown as green points. Significance is shown for considered benchmark scenarios as a function of $2m_{H^\pm}$, for 1.5 TeV (left) and 3 TeV (right) CLIC running. The red line shows 5σ threshold.

Table 6.3: Statistical significances obtained in the analysis, using full detector simulation, for every considered benchmark point. The results are shown together with the corresponding charged scalar masses and the cross sections for the scalar pair-production.

Benchmark point	1.5 TeV signif.	1.5 TeV σ [fb]	3 TeV signif.	3 TeV σ [fb]	$m_{H^\pm}$ [GeV]
on-shell W					
BP21	22.4	8.09	20.5	4.21	299.536
BP23	37.2	12.5	26.5	5.77	190.822
off-shell W^*					
HP3	4.2	0.629	14.1	1.78	633.48
HP17	9.8	2.44	6.9	1.69	440.624
HP20	13.3	1.32	17.2	1.51	543.794

from fast and full simulation seems to increase with decreasing mass difference, $m_{H^\pm} - m_H$, as visualized in Figure 6.5.

The reason for this discrepancy has to be not only the different level of detector response simulation (realistic vs. approximate), but also the influence of the overlay events. It can be seen on the distributions of the dijet mass and polar angle in Figures 6.1 and 6.2. Small mass difference between IDM scalars results in low mass of virtual $W^\pm$, and hence in the small energy accessible for its decay products. The hadrons originating from soft interactions contribute to the low-energy signal jets more significantly, than they would to the high-energetic ones, and smear their mass and energy distributions into higher values. This beam-induced background also "pushes" the jets towards smaller angles (with respect to the beam axis). All these effects make the signal events' characteristics more similar to the SM background (which is also influenced by the overlay), and more difficult to distinguish – both on the preselection and the MVA level. This is why, the decrease in statistical significance for scenarios with lower values of $m_{H^\pm} - m_H$ is clearly visible.

For the proper validation of fast simulation one also has to take into account that smaller

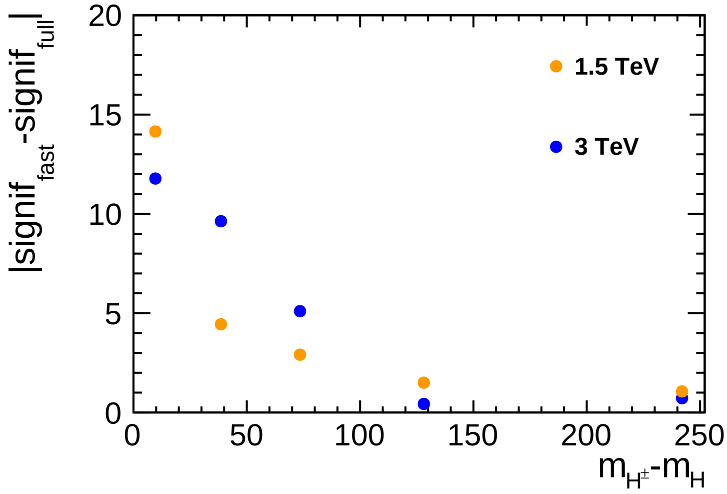

Figure 6.5: Absolute values of differences between resulting significances from fast and full simulation analysis. Orange points show values for 1.5 TeV and the blue points for 3 TeV.

number of background channels was considered in the full simulation analysis[1]. The impact of the additional background channels, also visible in Figures 6.1 and 6.2, is not large. For most points the significance obtained from the fast detector simulation analysis grows, when the same background channels as for the full simulation are considered and other channels are neglected. In particular, six-fermion channels $qq\ell\nu\ell\nu$ and $qq\ell\nu\nu\nu$, that significantly contributed to the total background in the fast simulation study (cf. Table 5.1) are important – even though the cross sections for these processes are small, the preselection cuts were unable to remove them from the sample. Presented results indicate that it is easier for the MVA to separate signal from background without the presence of those two background channels (and a few others as well). Furthermore, with the smaller number of samples and smaller numbers of generated events, the algorithm is much more likely to overtrain – especially for the signal, where only two to three samples are used in the training procedure. The overtraining can result in reduced significance (for the test sample), fortunately the effect is not large. On the other hand, the small number of signal samples allow BDT to tune to each particular scenario better (which should not be considered as overtraining). This again can have bigger impact for scenarios with small $m_{H^\pm} - m_H$, but seems to not affect the validation greatly as a whole. Nevertheless, it has to be kept in mind in the evaluation of full simulation results.

[1]Production of large background samples with the full detector simulation takes a lot of time, even when using grid resources. Some of the samples, which were included in the fast simulation, have not been processed yet.

Chapter 7

Conclusions

Many indications exist, that there is physics Beyond Standard Model yet to be discovered. As it has not been observed at the LHC yet, concepts of new, future colliders are being considered to offer complementary approaches to the current searches. For assessment of these projects, detailed studies are needed to investigate their scientific potential and facilitate the decision regarding their future construction. The studies are based on physics simulations, which allow to evaluate research potential of the given collider, its expected sensitivity for various BSM effects. Such an analysis was previously performed for the CLIC experiment concerning the discovery of the pair-produced Inert Doublet Model scalars in the fully-leptonic channel. Results of the generator-level study showed, that the observation of IDM particles for high-mass scenarios would not be possible with this final state.

In this work, the CLIC detector potential for discovering the IDM scalars was investigated in the semi-leptonic channel, as it provides much higher signal statistics. Detailed study with the realistic fast simulation of the detector response was performed using DELPHES package, and validated in the analysis based on the full simulation of the detector response with GEANT4 toolkit. In both cases, a cut-based preselection, followed by a multivariate analysis with Boosted Decision Trees was applied, which allowed to strongly reduce background contamination of the data. Results indicate, that heavy charged IDM particles could be observed at CLIC in the overwhelming majority of the considered scenarios, up to the scalar masses of about 1 TeV. High statistical significance, up to 60σ for low-mass scenarios, is expected in most cases, both at the 1.5 TeV and the 3 TeV CLIC running stages. Only for one of the 23 considered benchmark IDM scenarios, the expected deviations from Standard Model predictions do not exceed the 5σ discovery threshold at neither of the energy stages.

Subsequent full simulation analysis confirmed the fast simulation results. All of the five validation scenarios considered could be potentially observed at CLIC, also when taking the full detector response simulation into account. However, the influence of the beam-induced background (not included in the fast detector simulation) noticeably decreased the signal significance. Limited number of background samples accessible for the full simulation study could also lead to a slightly overestimated significance.

Presented study demonstrates, that the discovery of IDM charged scalar pair production at CLIC is possible, even with very high statistical significance, depending on the exact parameter values. It also shows, that semi-leptonic channel allows observation with much higher singificance, than in the search based on purely leptonic final state. Still, as leptons are reconstructed with much higher precision than the hadronic final state, the leptonic channel is much less affected by the backgrounds. Therefore, the leptonic signature might still be very useful in searching for scenarios with small differences in scalars' masses.

Appendix

Table A.1: Benchmark points considered in thesis, accessible at e^+e^- colliders with $\mathcal{O}$ (TeV) center-of-mass energies. $M_h = 125.1$ GeV for all points. BP21 and HP10 provide 100% dark matter relic density [33].

No.	M_H	M_A	$M_{H^\pm}$	Z on-shell	W on-shell	DM >50%	λ_2	λ_{345}	$\Omega_H h^2$
BP23	62.69	162.397	190.822	✓	✓	✓	2.63894	0.0056	0.064038
BP18	147	194.647	197.403				0.387	-0.018	0.0017718
BP21	57.475	288.031	299.536	✓	✓	✓	0.929911	0.00192	0.11946
HP1	176	291.36	311.96	✓	✓		1.4895	-0.1035	0.00072156
HP2	557	562.316	565.417			✓	4.0455	-0.1385	0.072092
HP3	560	616.32	633.48				3.3795	-0.0895	0.001129
HP4	571	676.534	682.54	✓	✓		1.98	-0.471	0.00056347
HP5	671	688.108	688.437				1.377	-0.1455	0.024471
HP6	713	716.444	723.045				2.88	0.2885	0.035152
HP7	807	813.369	818.001				3.6675	0.299	0.032393
HP8	933	939.968	943.787			✓	2.9745	-0.2435	0.09639
HP9	935	986.22	987.975				2.484	-0.5795	0.0027958
HP10	990	992.36	998.12			✓	3.3345	-0.051	0.12478
HP11	250.5	265.49	287.226				3.90814	-0.150071	0.00535
HP12	286.05	294.617	332.457				3.29239	0.112124	0.00277
HP13	336	353.264	360.568				2.48814	-0.106372	0.00937
HP14	326.55	331.938	381.773				0.0251327	-0.0626727	0.00356
HP15	357.6	399.998	402.568				2.06088	-0.237469	0.00346
HP16	387.75	406.118	413.464				0.816814	-0.208336	0.0116
HP17	430.95	433.226	440.624				3.00336	0.082991	0.0327
HP18	428.25	453.979	459.696				3.87044	-0.281168	0.00858
HP19	467.85	488.604	492.329				4.12177	-0.252036	0.0139
HP20	505.2	516.58	543.794				2.53841	-0.354	0.00887

Table A.2: Expected statistical significance for discovery of the charged IDM scalar pair-production at CLIC, as resulting from the fast simulation study for IDM benchmark points considered. Also shown are the production cross sections, σ, for the considered semi-leptonic channel and charged scalar masses, $m_{H^\pm}$. Four scenarios are beyond kinematic reach of the 1.5 TeV running stage.

Benchmark point	1.5 TeV signif.	1.5 TeV σ [fb]	3 TeV signif.	3 TeV σ [fb]	$m_{H^\pm}$ [GeV]
on-shell W					
BP21	21.4	8.19	16.6	4.27	299.536
BP23	36.1	12.6	22.4	5.68	190.822
HP1	26.8	7.59	19.8	4.01	311.96
HP4	0.5	0.23	6.9	1.59	682.54
off-shell W^*					
BP18	61.9	11.9	28.9	5.2	197.403
HP2	5.3	1.5	10.9	0.792	565.417
HP3	3.7	0.634	13.2	1.73	633.48
HP5	1.9	0.198	25.5	1.43	688.437
HP6	0.1	0.0372	13.7	1.07	723.045
HP7	-	-	13.3	1.01	818.001
HP8	-	-	8.0	0.693	943.787
HP9	-	-	11.0	0.607	987.975
HP10	-	-	2.8	0.432	998.12
HP11	59.4	7.23	38.9	3.51	287.226
HP12	35.7	3.78	23.6	2.09	332.457
HP13	53.8	5.74	46.1	3.15	360.568
HP14	21.1	2.24	14.7	1.40	381.773
HP15	44.8	4.77	38.2	2.99	402.568
HP16	44.3	4.34	44.5	2.74	413.464
HP17	16.1	2.45	21.3	1.42	440.624
HP18	36.5	3.31	42.2	2.48	459.696
HP19	28.6	2.59	38.9	2.28	492.329
HP20	17.4	1.33	29.7	1.54	543.794

Bibliography

[1] D. Sokolowska et al., *Inert Doublet Model signatures at future e^+e^- colliders*, 2019 European Physical Society Conference on High Energy Physics, 2019, arXiv: 1911.06254 [hep-ph].

[2] A. F. Zarnecki et al., *Searching Inert Scalars at Future e^+e^- Colliders*, International Workshop on Future Linear Colliders, 2020, arXiv: 2002.11716 [hep-ph].

[3] L. Linssen et al., eds., *Physics and Detectors at CLIC: CLIC Conceptual Design Report, CERN-2012-003*, CERN, 2012.

[4] P. Burrows et al., eds., *The Compact Linear Collider (CLIC) - 2018 Summary Report*, vol. 1802, CERN Yellow Reports: Monographs, 2018, DOI: 10.23731/CYRM-2018-002, URL: http://cds.cern.ch/record/2652188.

[5] CLIC input to the European Strategy for Particle Physics Update: https://clic.cern/european-strategy.

[6] CLIC, Compact Linear Collider Project, *CLIC accelerator schematic diagrams* (2019), General Photo, URL: https://cds.cern.ch/record/2655160.

[7] A. Robson et al., eds., *The Compact Linear e^+e^- Collider (CLIC): Accelerator and Detector*, 2018, arXiv: 1812.07987 [physics.acc-ph].

[8] *The Compact Linear Collider (CLIC) - Project Implementation Plan* (2018), ed. by M. Aicheler et al., DOI: 10.23731/CYRM-2018-004, arXiv: 1903.08655 [physics.acc-ph].

[9] *CLICdp Website*, https://clicdp.web.cern.ch/content/clic-detector, Accessed: 2020-04-13.

[10] D. Arominski et al., CLICdp, *A detector for CLIC: main parameters and performance* (2018), CLICdp-Note-2018-005, arXiv: 1812.07337 [physics.ins-det].

[11] J. Marshall, A. Münnich, M. Thomson, *Performance of Particle Flow Calorimetry at CLIC*, Nucl. Instrum. Meth. A **700** (2013) 153, DOI: 10.1016/j.nima.2012.10.038, arXiv: 1209.4039 [physics.ins-det].

[12] M. Thomson, *Particle flow calorimetry and the PandoraPFA algorithm*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **611** (2009) 25, ISSN: 0168-9002, DOI: 10.1016/j.nima.2009.09.009.

[13] A. F. Zarnecki, *On the physics potential of ILC and CLIC*, 19th Hellenic School and Workshops on Elementary Particle Physics and Gravity, 2020, arXiv: 2004.14628 [hep-ph].

[14] W. Kilian, T. Ohl, J. Reuter, *WHIZARD: Simulating Multi-Particle Processes at LHC and ILC*, Eur. Phys. J. C **71** (2011) 1742, DOI: 10.1140/epjc/s10052-011-1742-y, arXiv: 0708.4233 [hep-ph].

[15] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, DOI: 10.1016/j.cpc.2015.01.024, arXiv: 1410.3012 [hep-ph].

[16] S. Agostinelli et al., *Geant4 — a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002, DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.

[17] K. Bos et al., *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*, Technical Design Report LCG, CERN, Geneva, 2005, URL: http://cds.cern.ch/record/840543.

[18] J. de Favereau et al., DELPHES 3, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, JHEP **02** (2014) 057, DOI: 10.1007/JHEP02(2014)057, arXiv: 1307.6346 [hep-ex].

[19] E. Leogrande et al., *A DELPHES card for the CLIC detector* (2019), CLICdp-Note-2018-007, arXiv: 1909.12728 [hep-ex].

[20] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, http://www.deeplearningbook.org, MIT Press, 2016.

[21] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: physics/0703039 [physics.data-an].

[22] R. Brun, F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997), ed. by M. Werlen, D. Perret-Gallix 81, DOI: 10.1016/S0168-9002(97)00048-X.

[23] S. Chatrchyan et al., CMS, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30, DOI: 10.1016/j.physletb.2012.08.021, arXiv: 1207.7235 [hep-ex].

[24] G. Aad et al., ATLAS, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, DOI: 10.1016/j.physletb.2012.08.020, arXiv: 1207.7214 [hep-ex].

[25] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264, DOI: 10.1103/PhysRevLett.19.1264.

[26] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964), ed. by J. Taylor 508, DOI: 10.1103/PhysRevLett.13.508.

[27] Y. Akrami et al., Planck, *Planck 2018 results. I. Overview and the cosmological legacy of Planck* (2018), arXiv: 1807.06205 [astro-ph.CO].

[28] A. Einstein, *The Foundation of the General Theory of Relativity*, Annalen Phys. **49** (1916), ed. by J.-P. Hsu, D. Fine 769, DOI: 10.1002/andp.200590044.

[29] A. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Sov. Phys. Usp. **34** (1991) 392, DOI: 10.1070/PU1991v034n05ABEH002497.

[30] Y. Fukuda et al., Super-Kamiokande, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, DOI: 10.1103/PhysRevLett.81.1562, arXiv: hep-ex/9807003.

[31] N. G. Deshpande, E. Ma, *Pattern of Symmetry Breaking with Two Higgs Doublets*, Phys. Rev. D **18** (1978) 2574, DOI: 10.1103/PhysRevD.18.2574.

[32] A. Ilnicka, M. Krawczyk, T. Robens, *Inert Doublet Model in light of LHC Run I and astrophysical data*, Phys. Rev. D **93** (2016) 055026, DOI: 10.1103/PhysRevD.93.055026, arXiv: 1508.01671 [hep-ph].

[33] J. Kalinowski et al., *Benchmarking the Inert Doublet Model for e^+e^- colliders*, JHEP **12** (2018) 081, DOI: 10.1007/JHEP12(2018)081, arXiv: 1809.07712 [hep-ph].

[34] A. Albert et al., Fermi-LAT, DES, *Searching for Dark Matter Annihilation in Recently Discovered Milky Way Satellites with Fermi-LAT*, Astrophys. J. **834** (2017) 110, DOI: 10.3847/1538-4357/834/2/110, arXiv: 1611.03184 [astro-ph.HE].

[35] E. Aprile et al., XENON, *Dark Matter Search Results from a One Ton-Year Exposure of XENON1T*, Phys. Rev. Lett. **121** (2018) 111302, DOI: 10.1103/PhysRevLett.121.111302, arXiv: 1805.12562 [astro-ph.CO].

[36] J. Kalinowski et al., *Exploring Inert Scalars at CLIC*, JHEP **07** (2019) 053, DOI: 10.1007/JHEP07(2019)053, arXiv: 1811.06952 [hep-ph].

[37] R. Franceschini et al., *The CLIC Potential for New Physics*, **3/2018** (2018), ed. by J. de Blas, DOI: 10.23731/CYRM-2018-003, arXiv: 1812.02093 [hep-ph].

[38] S. Ovyn, X. Rouby, V. Lemaitre, *DELPHES, a framework for fast simulation of a generic collider experiment* (2009), arXiv: 0903.2225 [hep-ph].

[39] M. A. Weber, *Jet Performance at CLIC* (2018), CLICdp-Note-2018-004, URL: https://cds.cern.ch/record/2648827.

[40] M. Boronat et al., *Jet reconstruction at high-energy electron–positron colliders*, Eur. Phys. J. C **78** (2018) 144, DOI: 10.1140/epjc/s10052-018-5594-6, arXiv: 1607.05039 [hep-ex].

[41] M. Cacciari, G. P. Salam, G. Soyez, *FastJet User Manual*, Eur. Phys. J. C **72** (2012) 1896, DOI: 10.1140/epjc/s10052-012-1896-2, arXiv: 1111.6097 [hep-ph].

[42] M. Frank et al., *DD4hep: A Detector Description Toolkit for High Energy Physics Experiments*, Journal of Physics: Conference Series **513** (2014) 022010, DOI: 10.1088/1742-6596/513/2/022010.

[43] F. Gaede, J. Engels, *Marlin et al - A Software Framework for ILC detector R&D*, EUDET Report (2007),
URL: https://www.eudet.org/e26/e27/e584/eudet-report-2007-11.pdf.

[44] A. A. Maier, *CLIC_ analysis*, https://gitlab.cern.ch/amaier/CLIC_analysis/-/tree/master/processors/NtupleMaker, 2019.

[45] T. Charles et al., CLICdp, CLIC,
The Compact Linear Collider (CLIC) - 2018 Summary Report,
2/**2018** (2018), ed. by P. Burrows et al., DOI: 10.23731/CYRM-2018-002,
arXiv: 1812.06018 [physics.acc-ph].