Towards an update of the ILD ZHH analysis

Bryan Bliewert[1,*], *Caterina* Vernieri[2], *Dimitris* Ntounis[2], *Jenny* List[1], *Julie Munch* Torndal[1,3], and *Junping* Tian[4]

[1]Deutsches Elektronen-Synchrotron DESY,
Notkestr. 85, 22607 Hamburg, Germany
[2]SLAC National Accelerator Laboratory,
2575 Sand Hill Road, Menlo Park, California, United States
[3]Department of Physics, Universität Hamburg,
Jungiusstraße 9, 20355 Hamburg, Germany
[4]International Center for Elementary Particle Physics (ICEPP), The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033

E-mail: bryan.bliewert@desy.de, caterina@slac.stanford.edu, dntounis@stanford.edu, jenny.list@desy.de, julie.munch.torndal@desy.de, tian@icepp.s.u-tokyo.ac.jp,

Abstract. The double Higgs-strahlungs process $e^+e^- \to ZHH$ allows to access the Higgs self-coupling at center-of-mass energies above 450 GeV. Its cross-section exhibits a very different behavior as a function of the value of the self-coupling than fusion-type processes like gluon-gluon fusion at LHC (and future hadron colliders) and WW / ZZ fusion at higher energy lepton colliders. Therefore it adds unique information to the picture, in particular should the value of the Higgs self-coupling differ from its Standard Model prediction. The last full evaluation of the potential of the ILC to measure this process is more than ten years old, and since then many of the reconstruction tools have received very significant improvements. This contribution presents the ongoing work in the ILD collaboration to update the ZHH projections for the next European Particle Physics Strategy Update.

This work was carried out in the framework of the ILD Concept Group
ILD-PHYS-PROC-2024-01

*Speaker

1 Introduction

Since the discovery of a Standard Model (SM)-like Higgs boson at the LHC in 2012 [1, 2], measuring the value of the Higgs self-coupling $\lambda = \lambda_{hhh}$ to experimentally establish the mechanism of electroweak symmetry breaking is among the most important goals of particle physics.

While in a pure tree-level picture deviations of λ from its SM value are expected to be small if no deviations have been observed in other couplings, it has recently been shown [3] that this is not at all true once loop corrections are included: Already in extension of the SM Higgs sector as simple as a Two-Higgs-Doublet-Model (2HDM), λ can be driven to significantly larger values than in the SM by loop corrections with the additional Higgs bosons, *even if all other couplings are perfectly SM-like*. Thus it is of utmost importance to plan future colliders such that they are able to measure λ with the least possible model-dependence.

Double-Higgs production at the LHC and at e^+e^- colliders with $\sqrt{s} > 450\,\text{GeV}$ gives direct tree-level sensitivity to the self-coupling. As opposed to many other measurements, the experimentally achievable precision at any collider depends significantly on the actual value of λ realised in nature. A unique feature of the double Higgs-strahlungs process $e^+e^- \to ZHH$ is that its cross-section and the sensitivity to the value of λ increase when $\lambda > \lambda_{\text{SM}}$ [4]. This holds in particular in the region where λ is between 50% and a few times larger than the SM prediction, which is of special interest to models of electroweak baryogenesis. The cross-sections and the resulting sensitivity of vector–boson–fusion–like processes, including both gg fusion in pp collisions as well as WW fusion in e^+e^- collisions at higher center-of-mass energies, instead drops when $\lambda > \lambda_{\text{SM}}$. over a wide range of potential value of λ.

This is a strong motivation for future e^+e^- colliders which allow direct measurements of λ from double-Higgs production. One of them is the International Linear Collider (ILC) with the International Large Detector (ILD) detector concept. An energy stage of this machine at an energy of about 500 GeV would allow access to double-Higgs production, and in particular to double Higgs-strahlung $e^+e^- \to ZHH$. The last full analysis of this process was carried out ten years ago [5]. In light of the imminent input to the European Particle Physics Strategy Update (EPPSU) next year, there are now a renewed interest and a global effort to extract limits with state-of-the-art tools, benefiting both reconstruction and analysis. This effort aims to update the ILD projections for the ZHH process, and the extraction of λ based on detailed, Geant4-based detector simulation, including all relevant backgrounds. Also, following previous and promising studies [6], an additional analysis at a slightly higher center-of-mass energy of $\sqrt{s} = 550\,\text{GeV}$ will be carried out. Due to limited person power, the main focus will be on the $HH \to 4b$ channel, which covers 33.7% of all double-Higgs events. For the Z boson, the decays into jets, muons, electrons and neutrinos will be included, like in the previous analysis. Optionally the Z decays into τ leptons could be added.

The first section of this contribution presents improvements due to state-of-the-art tools developed during the last analysis, many of which are driven by machine-learning (ML). After that, pathways for further potential improvements are presented and a conclusion is given.

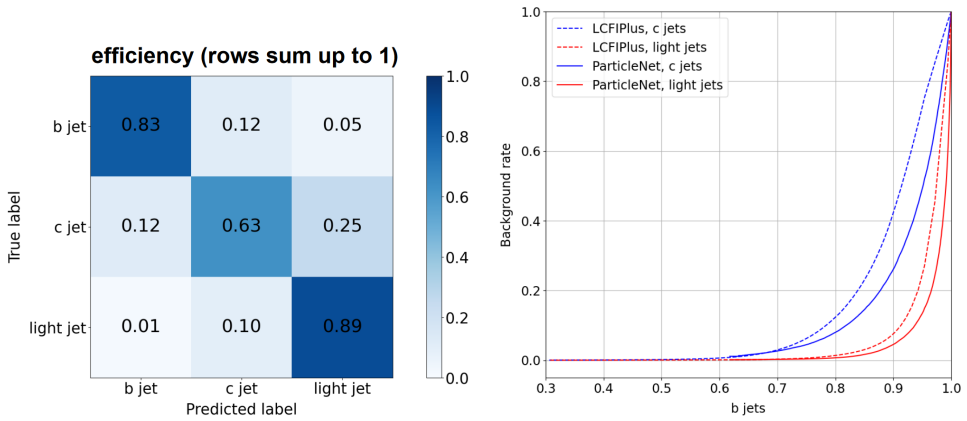
2 State-of-the-art Analysis Tools

Since the last full ZHH analysis, significant improvements could be made to reconstruction and analysis tools. The current benchmark for both jet clustering and tagging at ILD is LCFIPlus [7]. In the following, state-of-the-art methods for b-tagging, kinematic fitting and particle identification are presented.

2.1 Jet-tagging with ParticleNet

One crucial parameter in the ZHH analysis is the b-tagging efficiency. In the $4b$-channel, improvements of the b-tagging efficiency (at the same background rejection rates) enter the analysis by forth power.

In recent years, machine learning could be applied very successfully [8, 9] to the task of jet-flavour tagging. One example is the ParticleNet architecture, which treats jets as point clouds and is thus permutation invariant. It is able to process feature data of both jet constituents (i.e. particle flow objects, PFOs) and secondary vertices. The latter is beneficial for the tagging of heavy quark jets such as b-jets. ParticleNet has been implemented [10] in the Marlin framework and trained using fully-simulated physics samples with 6 jets at ILD with a center-of-mass energy of $\sqrt{s} = 500\,\text{GeV}$. A confusion matrix including b, c and light (u, d, s) quarks as well as a ROC-curve comparing the ParticleNet model with LCFIPlus are shown in Fig. 1.



(a) Jet-tagging efficiency

(b) ROC-curve

Figure 1: Jet-tagging performance with ParticleNet at ILD and ILC500. (a) Tagging efficiencies for separating b- from c- and light jets at an example working point. (b) Performance comparison of b- vs. c- and light-jet tagging by ROC curves with the benchmark algorithm LCFIPlus. The plot shows background efficiency vs. signal efficiency (the closer to the bottom-right corner, the better). From [10].

As can be seen in Fig. 1a, a b-tagging efficiency ϵ_b of around 83% could be obtained by using the ParticleNet model at an example working point with a mis-tag rate of 12% to falsely identify a c jet as b jet. Comparing the overall performance of the ML-based tagger with that of LCFIPlus as in Fig. 1b shows a consistently superior background suppression

when identifying b-jets. For the separation of b jets with light jets (c jets), an up to 4% increase in efficiency over the already improved LCFIPlus [11] can be gained at the same purities, which are around 99% (88%). Earlier studies [5] have shown that an up to 11% relative improvement on the precision of the self-coupling can be expected from a 5% relative increase in ϵ_b. The results in Fig. 1 show that this expected improvement is clearly within reach.

2.2 Kinematic fitting

Another important ingredient to the previous ZHH analysis was kinematic fitting. At the time of the previous analysis, the kinematic fit already was able to account for initial-state photon radiation [12], but the correction for missing energy from semi-leptonic heavy flavour decays inside the b jets as well as the overall estimate of the measurement errors on the jet energy and the jet angles were very rudimentary. In both areas, significant progress has been achieved in the last years [13–15], leading to a massive improvement of the kinematic reconstruction of $H \to b\bar{b}$ decays, illustrated in Fig. 2 for the somewhat simpler case of $ZH \to \mu^+\mu^- b\bar{b}$ at $\sqrt{s} = 250\,\text{GeV}$.

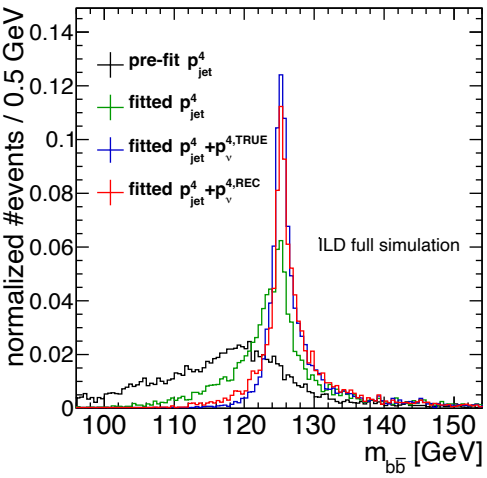


Figure 2: Reconstructed di-jet masses from $H \to b\bar{b}$ decays before and after kinematic fitting: no kinematic fit, no neutrino correction (black), kinematic fitting without neutrino correction (green), kinematic fitting with neutrino correction using fully reconstructed (red) and cheated, i.e. truth-level (blue) neutrino momenta. From [14].

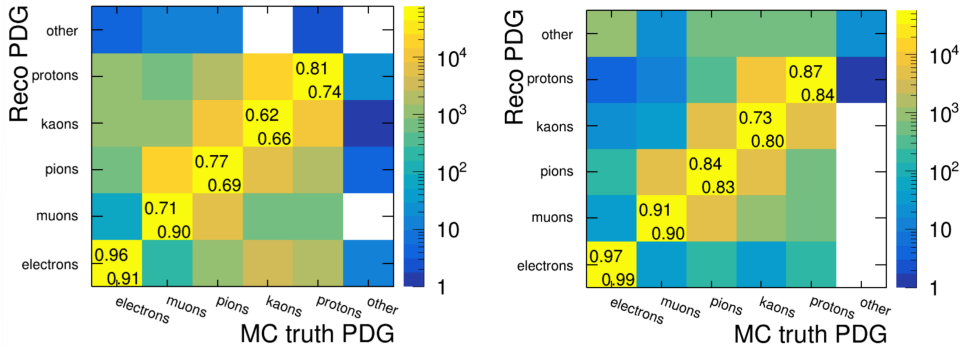
Transferring these developments to the more complex case of ZHH events is work-in-progress, which is expected to help both with identifying the correct di-jet pairs originating from one boson as well as to separate ZHH events from the most challenging backgrounds like ZZH production.

2.3 Comprehensive Particle Identification (CPID)

The ZHH analysis profits further from improvements in particle identification (PID) in multiple ways. One example is the leptonic signal channel. There, after an isolated lepton criterion selects charged leptons, only leptons of the same family and with opposite charge

are included in the pairing to reconstruct the leptonic Z decay. Because two leptons are paired, improvements to the PID efficiency of electrons and muons go into the analysis squared.

As a PID framework, CPID [16] has been designed from the ground-up towards modularity for both different inference modules as well as data sources and detector systems. These range over dE/dx, cluster shape and time-of-flight (TOF). The latter has so far not been included in the benchmark PID solution ("LikelihoodPI"). An example of the performance gain when using CPID at ILD is illustrated in Fig. 3.



(a) Likelihood PID

(b) CPID

Figure 3: Confusion matrices showing the performance of (a) the benchmark PID algorithm and (b) the new CPID (right). On the diagonal elements, efficiency and purity are given in the top left and bottom right, respectively. From [16].

The confusion matrix for CPID in Fig. 3b shows excellent separation for leptons and still very good results for separating pions, kaons and protons. Compared to the benchmark, the PID efficiency is significantly improved by a relative 10-25% for all considered particles except electrons. For all particles except the muons, also the purity is improved by a relative 15-20% at the same time.

3 Future Analysis Tools

Additional to the state-of-the-art and production-ready improvements described earlier, two challenges for future research shall be presented here with proof-of-concept solutions. They address the tasks of jet clustering and event classification.

3.1 More accurate jet clustering with ML

Previous studies [5] have shown that the jet clustering is one of the leading sources of error in the ZHH analysis, impacting the sensitivity to λ by approximately a factor of two. This is due to the interplay of two effects: mis-clustering and mis-pairing of jets to di-jets. They can be illustrated by making use of a unique feature of the event data model [17], namely mappings between particles and di-jets from reconstruction and their corresponding counterpart from full simulation. Using these mappings of PFOs to di-jets and matching each reconstructed

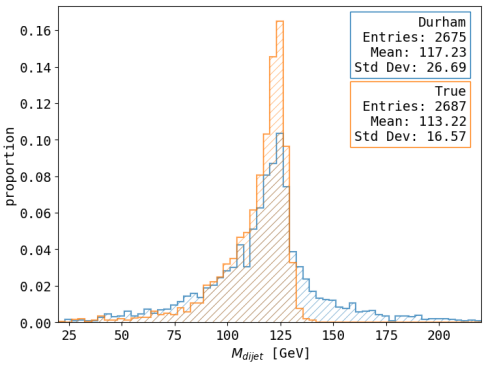


Figure 4: Reconstructed di-jet masses with the Durham algorithm and cheated ("True") using TrueJet [18].

di-jet to it's "true counterpart" by angular overlap, a "true" di-jet can be defined and its associated mass calculated.

This is shown for hadronically decaying Higgs bosons in ZHH events in Fig. 4. The much sharper mass distribution for the "true" di-jets shows that a more accurate jet clustering algorithm would significantly benefit any downstream analysis that uses the reconstructed di-jet mass. For the Durham case, the jets are paired to two di-jets using the best χ^2 matching assuming the Higgs mass.

A more thorough analysis is possible by quantifying the purity π_E and efficiency ϵ_E of each di-jet's energy [6]. When imposing a 95% boundary on both axes, all di-jets can be grouped into four categories as given in Table 1.

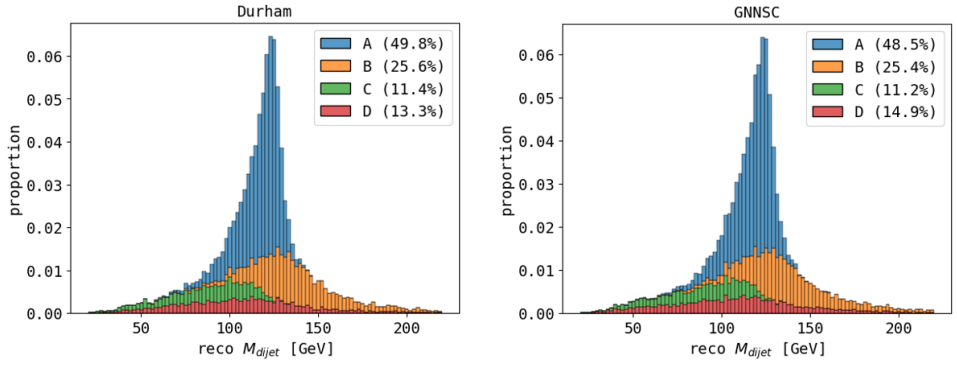
Category	$\pi_E = E_{corr}/E_{true}$	$\epsilon_E = E_{corr}/E_{reco}$
A	$\geq 95\%$	$\geq 95\%$
B	$\geq 95\%$	$< 95\%$
C	$< 95\%$	$\geq 95\%$
D	$< 95\%$	$< 95\%$

Table 1: Categorization of reconstructed di-jets by energy fraction. E_{corr} is the energy of the correctly clustered PFOs (i.e. union of PFOs present in the reconstructed and associated true di-jet).

Assuming the categorization given in Table 1, the higher the fraction of di-jets in category A and the lower in the other fractions, the better the average reconstruction performance.

One possible approach towards more accurate jet clustering algorithms is to use ML, precisely graph neural networks (GNNs) to calculate scores for whether or not jet constituents shall be clustered together. Following this idea, a hybrid model ("GNNSC") has been developed as a proof-of-concept study. This model uses the TransformerConv operator [19] sandwiched between linear input layers and a Tanh activation layer to calculate scores, and spectral clustering (SC) on the affinity matrix $A \in \mathbb{R}^{n \times n}$ to create m clusters for n PFOs. The scoring module takes in only the four-momenta of the PFOs. It is trained

fully-supervised on ZHH events and using labels from the true jet to PFO-matching described above (ignoring gluon splittings), while the spectral clustering requires no training. Results for reconstructed di-jet masses using Durham and GNNSC are shown in Fig. 5.



(a) Durham (b) GNNSC

Figure 5: Di-jet masses for hadronically decaying Higgs bosons in ZHH samples at 500 GeV as reconstructed by (a) the Durham algorithm and (b) the experimental GNNSC model.

The Higgs peak at 125 GeV is well reconstructed using either Durham or the GNNSC method, as evident from Fig. 5. More precisely, both the full-width at half-maximum (FWHM) of the two distributions and the distribution of the mis-clustering categories A-D over the di-jets are effectively indistinguishable. Further studies have shown that the model successfully generalizes to other physics processes (e.g. ZZH) well. This leads us to the conclusion that already with reasonably complicated ML architectures, it is possible to reach the performance of current benchmark jet clustering algorithms. More refined architectures with inbuilt physical symmetries and properties (IRC-safety) as well as fully-differentiable models could allow additional gains and allow to surpass the performance of current jet clustering algorithms at future Higgs factories.

3.2 Event classification with Matrix Elements

An important challenge in the ZHH analysis is the separation of irreducible backgrounds, for example ZZH events. Example Feynman diagrams for both processes in the leptonic signal channel (here $Z \to \mu^+\mu^-$) are given in Fig. 6.

Figure 6: Some of the Feynman diagrams for (a) the ZHH and (b) the ZZH process.

Detector effects as well as missing momentum from semi-leptonic decays of B-mesons and other particles complicate the separation of ZHH and ZZH events when observables such as the invariant mass are used.

Another observable for separating these processes and using all available kinematic information is the hard-scattering matrix element. It can be calculated for both processes and including the remaining contributing Feynman diagrams with the reconstructed kinematics of the final state leptons and b-jets. The ratio $r = M_{ZHH}/M_{ZZH}$ of the matrix elements can then be used to construct a discriminator. An example for a sample of ZHH and ZZH events is given in Fig. 7.

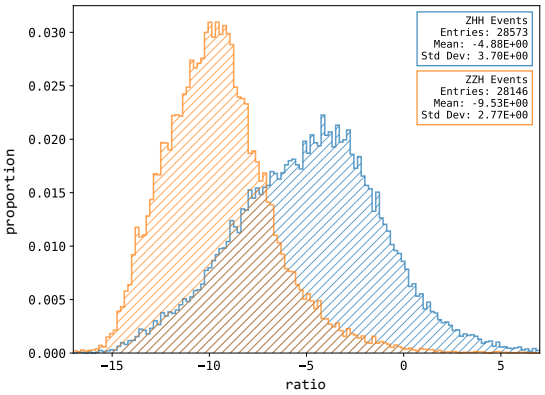


Figure 7: Ratio of matrix elements for the ZHH and ZZH hypothesis.

The histogram in Fig. 7 shows that the hard-scattering matrix elements calculated with the reconstructed kinematics already provide much separation power. Here, the kinematics of the Higgs boson are reconstructed by adding the four momenta of the associated jets and for the ZZH hypothesis, the average over the possible permutations is calculated.

However, to account for detector effects and parton showering, the approach must be extended. This procedure is known as the matrix element method (MEM) and described in literature [20, 21]. It allows to calculate the likelihood for an event to have happened under a certain hypothesis and is based on first principles. Instead of an evaluation at the measured kinematics, it requires a phase space integration over the matrix elements, functions describing the transfer of parton to jet kinematics and other features. As a proof-of-concept, this procedure has been carried with some simplifications, for example assuming perfect detector acceptance as well as efficiency and perfect measurement of the muon kinematics. Also, initial state radiation (ISR) is not accounted for. An on-shell condition has been assumed for the b quarks, such that the jet momentum transfer can be parameterized in spherical coordinates and using the energy and fixed mass. The transfer of jet angles and energy is manually fitted using Gaussian functions and the mappings between generator- and reconstructed level quanta (ignoring gluon splittings). Finally, the integral is solved using the VEGAS+ algorithm [22]. In accordance to the Neyman-Pearsson lemma, the two hypotheses can then be separated via the likelihood-ratio. The performance of the resulting discriminator is shown with a ROC-curve in Fig. 8.

The discriminator shows acceptable performance in Fig. 8 with an area-under-curve (AUC) of 0.73. In the future, improvements could be made by modeling the detector transfer

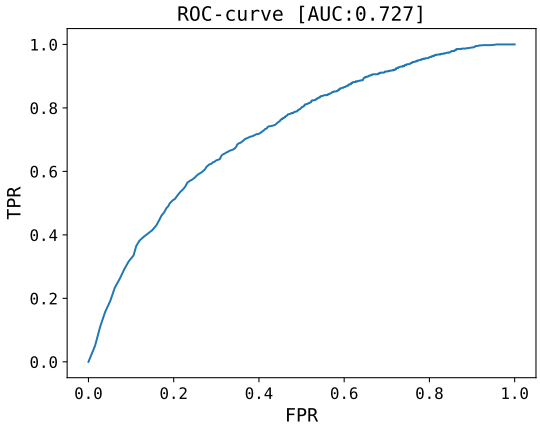


Figure 8: ROC-curve of the ZHH/ZZH discriminator using the full MEM procedure in terms of the true positive rate (TPR), i.e. signal efficiency, vs. false positive rate (FPR), i.e. background efficiency.

functions more accurately (e.g. using normalizing flows) and developing dedicated phase space parameterizations for each process.

4 Conclusion

About ten years ago, the expected precision on the Higgs self-coupling at ILC at 500 GeV has been evaluated based on detailed, Geant4-based simulation of the ILD detector concept. Based on the ZHH process with $Z \to q\bar{q}, e^+e^-, \mu^+\mu^-, \nu\bar{\nu}$ and $HH \to b\bar{b}b\bar{b}/b\bar{b}WW^*$, and assuming the SM value of the self-coupling, the projection achieved a precision of 27%.

The ongoing update of the ZHH analysis will benefit from a variety of production-ready improvements, especially in jet tagging, PID and using kinematic fitting. We expect these to increase the sensitivity $\Delta\lambda_{SM}/\lambda_{SM}$ from 27% to better than 20%. More improvements are possible in the removal of the $\gamma\gamma$-overlay (up to relative 15%) and by including $Z \to \tau^+\tau^-$ in the leptonic channel (up to relative 8%). This does neither include further improvements from a slightly higher $\sqrt{s} = 550\,\text{GeV}$ nor from the inclusion of the WW-fusion production, which are both under study as well.

Further future analysis improvements could be accomplished by developing more accurate jet clustering and event classification algorithms. Machine learning offers many opportunities for these tasks and can profit from more physical observables as well.

5 Acknowledgments

We would like to thank the LCC generator working group and the ILD software working group for providing the simulation and reconstruction tools and producing the Monte Carlo samples used in this study. This work has benefited from computing services provided by the ILC Virtual Organization, supported by the national resource providers of the EGI Federation and the Open Science GRID. In this study we widely used the National Analysis

Facility (NAF) [23] and the Grid computational resources operated at Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany. We thankfully acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2121 "Quantum Universe" 390833306.

References

[1] G. Aad et al. (ATLAS), Phys. Lett. B **716**, 1 (2012), 1207.7214

[2] P.C. Bhat (CMS), Nucl. Phys. B Proc. Suppl. **234**, 7 (2013)

[3] F. Arco, S. Heinemeyer, M.J. Herrero, Eur. Phys. J. C Part. Fields **82** (2022)

[4] J. Alison et al., Rev. Phys. **5**, 100045 (2020), 1910.00012

[5] C.F. Duerig, *Measuring the Higgs self-coupling at the international linear collider* (Deutsches Elektronen-Synchrotron, DESY, Hamburg, 2016), https://bib-pubdb1.desy.de/record/310520

[6] J. Torndal, J. List, D. Ntounis, C. Vernieri, *Higgs self-coupling measurement at future e+e- colliders*, in *Proceedings of The European Physical Society Conference on High Energy Physics — PoS(EPS-HEP2023)* (Sissa Medialab, Trieste, Italy, 2024)

[7] T. Suehara, T. Tanabe, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **808**, 109 (2016)

[8] H. Qu, L. Gouskos, Phys. Rev. D **101**, 056019 (2020)

[9] H. Qu, C. Li, S. Qian, arXiv [hep-ph] (2022)

[10] M. Meyer (2023), presented at the Second ECFA Workshop on e+e- Higgs/EW/Top Factories in Paestum (Salerno), https://agenda.infn.it/event/34841/contributions/207748/attachments/111477/159079/ECFA_Paestum.pdf

[11] T. Suehara, T. Tanabe, M. Kurata, J. Strube (2017), presentation at ILD software and technical meeting at the Institut de Physique Nucléaire de Lyon, https://agenda.linearcollider.org/event/7520/contributions/38829/attachments/31495/47441/170425-lcfiplus-ildsoftware-suehara.pdf

[12] M. Beckmann, B. List, J. List, Nucl. Instrum. Meth. A **624**, 184 (2010), 1006.0436

[13] U. Einhaus, B. Dudar, J. List, Y. Radkhorrami, J. Torndal, PoS **ICHEP2022**, 538 (2022), 2212.07264

[14] Y. Radkhorrami (2023), presented at the ILD Software and Analysis meeting, https://agenda.linearcollider.org/event/10079/contributions/52747/attachments/38973/61392/Yasser_Radkhorrami_ILDSWANA_20230621.pdf

[15] Y. Radkhorrami (2025), phD thesis in preparation

[16] U. Einhaus, *CPID: A comprehensive particle identification framework for future e^+e^- colliders* (2023), https://arxiv.org/abs/2307.15635

[17] F. Gaede, T. Behnke, N. Graf, T. Johnson, *Lcio - a persistency framework for linear collider simulation studies* (2003), physics/0306114, https://arxiv.org/abs/physics/0306114

[18] M. Berggren (2018), presented at the ILDsw meeting, KEK, https://agenda.linearcollider.org/event/7839/contributions/40940/attachments/32721/49770/berggren-ild-sw-kek-feb18-true.pdf

[19] Y. Shi, Z. Huang, S. Feng, H. Zhong, W. Wang, Y. Sun, *Masked label prediction: Unified message passing model for semi-supervised classification* (2021), 2009.03509, https://arxiv.org/abs/2009.03509

[20] K. Kondo, Journal of the Physical Society of Japan **57**, 4126 (1988)

[21] I. Volobouev, *Matrix element method in hep: Transfer functions, efficiencies, and likelihood normalization* (2011), 1101.2259
[22] G.P. Lepage, Journal of Computational Physics **439**, 110386 (2021)
[23] A. Haupt, Y. Kemp, J. Phys. Conf. Ser. **219**, 052007 (2010)