From Jets to Pixels: Flavour Tagging and Monolithic Pixel Sensors at the FCC-ee

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Eduardo V. Ploerer

aus

Österreich

Promotionskommission

Prof. Dr. Florencia Canelli (Vorsitz)

Prof. Dr. Freya Blekman

Prof. Dr. Michael Tytgat

Prof. Dr. Gino Isidori

Dr. Anna Macchiolo

Zürich, 2025

From Jets to Pixels: Flavour Tagging and Monolithic Pixel Sensors at the FCC-ee

---

A thesis

submitted in fulfilment of the requirements for the degree of Doctor of Sciences

in the

Faculty of Sciences & Bio-Engineering Sciences

of the

Vrije Universiteit Brussel

by

Eduardo V. Ploerer

from

Austria

Promotors

Prof. Dr. Michael Tytgat

Prof. Dr. Florencia Canelli

Supervisor

Prof. Dr. Freya Blekman

Brussels, 2025

DEUTSCHES ELEKTRONEN-SYNCHROTRON
Ein Forschungszentrum der Helmholtz-Gemeinschaft

DESY-THESIS-2025-016
September 2025

From Jets to Pixels: Flavour Tagging and Monolithic Pixel Sensors at the FCC-ee

by

E. Plörer

ISSN 1435-8085

NOTKESTRASSE 85 – 22607 HAMBURG

The work presented in this thesis has in part or wholly been supervised by:

Prof. Dr. Freya Blekman (DESY, Universität Hamburg)
Prof. Dr. Florencia Canelli (Universität Zürich)
Prof. Dr. Michael Tytgat (IIHE, Vrije Universiteit Brussel)
Dr. Anna Macchiolo (Universität Zürich)
Dr. Armin Ilg (Universität Zürich)

Abstract

The FCC-ee is the first stage of a two stage project which would envision a 92 km tunnel being used for an electron-positron collider experiment starting around 2045, and a proton-proton collider in the 2070's. The FCC-ee offers unprecedented precision tests of the Standard Model, owing to its clean leptonic collision environment and exquisite luminosity. The precise identification of decay products at future colliders is instrumental to the exploitation of the full physics potential. In this thesis two facets of this problem are explored: jet flavour tagging and charged particle tracking.

Jet flavour tagging describes the algorithmic identification of the initiating parton from hadronic decays. The prospect of identifying strange quark jets has emerged as a promising avenue to study a multitude of largely unexplored processes, including $\mathrm{Z} \rightarrow s\bar{s}$ production and rare Higgs boson decays. *DeepJetTransformer* is a transformer-based multiclassifier neural network developed by the CMS jet tagging team at the VUB, achieving state-of-the-art performance while being relatively lightweight. This thesis combines DeepJetTransformer with secondary vertexing and a novel implementation of K-short reconstruction at the FCC-ee to identify strange jets. Through the inclusion of different levels of $\mathrm{K}^{\pm}/\pi^{\pm}$ discrimination, strange tagging efficiencies ranging from 31.6% to 57.8% were obtained at a u, d jet efficiency of 10%, highlighting the need for charged Kaon discrimination at future colliders.

Monolithic Active Pixel Sensors (MAPS) combine the sensing node and readout circuitry into the same substrate, thus offering several advantages with respect to their hybrid counterparts. The Circuit Exploratoire 65 nm (CE-65), and its evolution CE-65v2, are MAPS test structures produced in the 65 nm TPSCo CMOS process to explore charge collection properties for a variety of configurations, including variations in pixel pitch, process modification, amplification scheme, and matrix geometry. In this thesis the lab characterisation of the CE-65v1 and CE-65v2 chips is reported, where charge collection efficiencies of 96% were achieved for all variants. In a subsequent beam test at CERN SPS a sub 3 µm spatial resolution was obtained for Standard process variants, satisfying FCC-ee requirements. The characterisation of the CE-65 family of chips has supplemented the APTS and DPTS studies in the validation of the 65 nm TPSCo process as a candidate technology for advanced particle detection applications, directly informing the development of future tracking detectors.

“The limits of my language mean the limits of my world.”
– Ludwig Wittgenstein

Acknowledgements

And for the next 170 or so pages, also your world. As with most journeys in life, it is seldom the destination, but the company one has that defines it. This is the only part of this thesis in which I have taken the liberty of writing in the first person, and will thus make full use of it.

These Acknowledgements begin in much the same way as my PhD began: with Freya. Freya has guided my work throughout, providing both direction and encouragement. While the official duty of my supervision shifted from Freya onto Jorgen, and finally onto Michael, I am grateful to both for their support and for helping me navigate the many bureaucratic burdens along the way. Throughout all stages of my PhD, it was Freya that provided the steady guidance that carried my research from its inception to its completion.

My first excursion into particle physics occurred during my Masters under the supervision of Florencia. I was grateful then, just as I am now, to have been able to work on such exciting topics, and I thank her also for her continued support during my PhD. I would like to thank Anna and Armin for their close supervision during my hardware work. Armin in particular was invaluable: from providing near-daily feedback, to accompanying me to test-beam campaigns, and commenting on virtually every chapter of this thesis.

I would like to thank the current and former members IPHC Strasbourg with whom I collaborated closely during the characterisation of the CE-65 test structures. In particular Serhiy, Ajit, and Szymon, whose expertise I made use of on several occasions.

I am thankful to David d'Enterria for his guidance and many helpful discussions.

Over the course of my PhD I had the great pleasure of being based at the IIHE, UZH, and DESY. The incomparable atmosphere of the IIHE was created by current and former members including Soumya, AR, Denise, Alexandre, Felix, Nordin, Jas, Hugues, Katka, Max, and Marta, as well as some of the members I have met more recently, including Juhee, Pavlo, Donya, Saranya, and Stef. I can rest assured that the superb atmosphere of the IIHE is alive and well. At UZH I was fortunate enough to share an office, and many enjoyable discussions, with Eslam. I also owe thanks to Kyle, who patiently entertained questions far removed from his own research, but whose answers always turned out surprisingly productive. Fabian, Weijie, Sascha, and Pascal were always good company when I found myself in the upstairs office. I would also like to thank the DESY CMS group for their welcoming atmosphere, in particular Lucia, Keila, and Gabriele.

Throughout the entirety of my PhD I was accompanied by my "academic twin" and dear friend, Kunal. From our first days in Brussels through to the final stretch of writing, we even managed to line up our thesis submission dates. Kunal has always been there through thick and thin, always ready to share in frustrations and celebrate every victory.

I would also be remiss to not mention my roommates at HJ who provided a constant source of company and levity during long months of work. Since the very beginning of my journey into physics, I have been accompanied by my dear friend Antonio, who has always been there to share a laugh, even when his ping pong abilities left much to be desired.

Finally, I am eternally indebted to my family, whose love and support has carried me through this journey. From my mother Isabel's endless affection, to my father Vinzenz's guidance and constant faith, to my sisters Victoria and Emilia, whose growth I have watched with pride. And of course Marijn, my bastion of strength, who ensured that the latter part of this journey was not only endurable, but joyful. There was never a moment in which my confidence truly wavered, for I knew that I carried the trust of those who believed in me.

Yours sincerely,

Eduardo

Author Contribution Statement

As the work presented in this thesis is collaborative in nature, and was performed together with a number of researchers across several institutes, this statement serves to highlight some of the more significant contributions the author made.

Jet Flavour Tagging

The DeepJetTransformer network was originally developed by the VUB CMS group [1], primarily in the context of heavy flavour tagging. The author adapted the network to the FCC-ee, including modifications necessary to the input dataset and network structure. All training and inference associated with the flavour tagging results in Chapter 6, including the feature importance studies, were performed by the author. The flavour tagging studies herein presented are summarised in Ref. [2], to which the author contributed extensively through model training and evaluation, interpretation of results, and manuscript preparation.

Laboratory Characterisation

The laboratory measurements presented in this thesis, particularly the ^{55}Fe source measurements, were performed by the author at the University of Zurich. The author significantly contributed to the DAQ software, implementing features such as thresholding to reduce data rates by a factor ~ 12, and adding compatibility for the 25 µm CE-65v1 chip. In parallel, the author developed portions of the analysis software and was solely responsible for developing the fitting routines used for the CE-65 laboratory characterisation. The analysis of all lab measurements presented in this thesis was performed by the author. Additional contributions include the automation of more tedious parts of data taking, including the biasing voltage scans presented in Section 9.4.

Testbeam Analysis

The author actively participated in the CE-65v2 testbeam campaigns at SPS (2024) and DESY (2023). The author developed a variety of supporting software and documentation for the CE-65v2 testbeam analysis, including containerisation of the analysis environment, to standardise and streamline the analysis workflow. Moreover, the author added support for the CE-65v2 staggered matrix geometry in the Corryvreckan framework [3], involving the implementation of the detector geometry and adjustments to the clusterisation algorithm.

The author is responsible for producing all results presented in Chapter 10, with the sole exception of the in-pixel resolution/efficiency of the 15 µm Modified with Gap process CE-65v2 chip. Preliminary results of the CE-65v2 laboratory characterisation and SPS 2024 testbeam were summarised in Ref. [4], which are drawn largely from the author's work. The author also prepared the initial manuscript draft.

Contents

1 Introduction **1**

I Theoretical and Experimental Foundations **5**

2 Standard Model **7**
- 2.1 Beyond the Standard Model 10

3 The Future Circular Collider Project **13**
- 3.1 Introduction to Collider Physics 13
- 3.2 The FCC Integrated Project 14
- 3.3 FCC-ee Parameters 15
- 3.4 Physics Programme 17
- 3.5 Detector Concepts 19
  - 3.5.1 IDEA 20
  - 3.5.2 CLD 21
  - 3.5.3 ALLEGRO 22
  - 3.5.4 ILD 23

4 Reconstruction **25**
- 4.1 Tracks 25
  - 4.1.1 Vertexing 26
- 4.2 Particle Reconstruction 27
  - 4.2.1 Leptons and Photons 27
  - 4.2.2 Hadrons 28
- 4.3 Jets 28

II Jet Flavour Tagging at the FCC-ee **33**

5 Machine Learning **35**
- 5.1 Neural Networks 36
- 5.2 Learning as Optimisation 39
- 5.3 Beyond Fully Connected Networks 45
- 5.4 Transformers and the (Self-) Attention Mechanism 48
- 5.5 Machine Learning in Jet Flavour Tagging 51

6 Jet Flavour Tagging at the FCC-ee with a Transformer-based Neural Network **53**
- 6.1 Jets 53
- 6.2 Input Features 54
- 6.3 Network Architecture 57

6.4 Training Details 60
6.5 Classification Results 60
    6.5.1 Dependence on Particle Identification 62
6.6 Feature Importance 65
6.7 Z Peak Extraction 67
6.8 Conclusions 70

III Monolithic Active Pixel Sensors for Future Colliders 73

7 Silicon Detectors 75
7.1 Semiconductors 75
7.2 Signal Formation 80
    7.2.1 Particle Interactions with Matter 80
    7.2.2 Signal: Induced Current 83
7.3 Signal Readout 85
    7.3.1 Readout Electronics 85
    7.3.2 Noise 86
7.4 Monolithic Pixel Detectors 87

8 CE-65 Chip 91
8.1 Chip Details 91
8.2 Process Modifications 92
    8.2.1 Standard Process 93
    8.2.2 Modified Process 93
    8.2.3 Modified with Gap Process 93
8.3 Amplification Scheme 94
    8.3.1 Source-Follower 94
    8.3.2 DC Preamplifier 95
    8.3.3 AC Preamplifier 95
8.4 Chip Readout 96
8.5 Matrix Arrangement 97
8.6 Summary of V1 and V2 Design Differences 98

9 Laboratory Characterisation 99
9.1 Lab Setup 99
    9.1.1 Data Acquisition Setup 99
    9.1.2 Data Acquisition Code 100
9.2 Noise Characterisation 101
9.3 Methodology of Radioactive Source Characterisation 105
    9.3.1 Fe-55 Emission Lines 105
    9.3.2 Clusterisation 106
    9.3.3 Fitting Methodology 108
9.4 Gain Calibration 109
9.5 Charge Sharing 113
    9.5.1 Seed vs Neighbours 113
    9.5.2 Charge Collection Efficiency 114

9.5.3 Contributions of Varying Cluster Sizes . . . 115
9.6 Gain Uniformity . . . 116
9.6.1 Individual Pixel Spectra . . . 116
9.6.2 Energy Resolution . . . 117
9.6.3 Pixel-by-Pixel Calibration . . . 119
9.7 Conclusions . . . 124

10 Testbeam Analysis **127**
10.1 Methodology . . . 127
10.1.1 Performance Metrics . . . 128
10.1.2 Corryvreckan . . . 128
10.2 CERN SPS 2024 Testbeam . . . 132
10.2.1 Setup . . . 132
10.2.2 Analysis Strategy for SPS . . . 134
10.2.3 Hit Detection Efficiency . . . 137
10.2.4 Spatial Resolution . . . 140
10.2.5 Charge Sharing . . . 145
10.2.6 Ramifications of a Staggered Matrix Geometry . . . 148
10.3 DESY 2023 Testbeam . . . 155
10.3.1 Setup . . . 155
10.3.2 Results . . . 156
10.4 Conclusions . . . 162

IV Conclusion and Outlook **165**

11 Conclusion **167**

References **171**

1 Introduction

The Standard Model (SM) provides a consistent and predictive framework describing fundamental particles and their interactions. It has since the 1980s become the established theory of particle physics, and a sizeable departure from the theory is yet to be observed. In 2012 [5] its last, and arguably one of its most central elements, was observed: the Higgs boson. While this marked a triumph of the theory, it was also an expected one, and thus its presence alone did not clearly chart a course forward. The Standard Model has proven robust in the face of extensive experimental scrutiny, yet it is known that it does not provide a complete description of nature as evidenced by a variety of phenomena that remain unexplained, including the absence of gravity in the framework, the existence of dark matter, and the observed matter-antimatter asymmetry.

The particle physics community is thus at a crossroads, where the clear necessity for studying the Higgs boson in detail has arisen. The most suitable experiment for such an endeavour is an electron collider, which owing to its clean experimental conditions is often dubbed a "precision machine". Electrons, unlike protons which are currently collided at the Large Hadron Collider (LHC) at CERN, are fundamental particles. Electron-positron collisions thus involve well-defined initial states, which are additionally not subject to the strong force. As a result, electron-positron collisions are considerably cleaner than their hadronic counterparts, with fewer final-state particles and a sizeably reduced background from particles not associated with the primary hard scattering. By operating near the Higgsstrahlung threshold, lepton colliders can produce a Higgs boson in association with the well-studied Z boson and *nothing else*[1], thus enabling the study of one of nature's most interesting objects. Beyond Higgs studies, lepton colliders enable a broad physics programme encompassing electroweak and strong interactions. This includes precision measurements of Z and W boson properties, which mediate the weak force. Precision studies of the strong force, including a per-mille determination of the strong coupling α_s, are key to global electroweak fits and other precision measurements where it might otherwise dominate as a parametric uncertainty.

In each of these a departure from SM predictions would imply the existence of a deeper underlying theory, that could be entering these measurements indirectly. Such an effect has been observed in the past, for instance with the top quark, where its loop contribution to $e^+e^- \to f\bar{f}$ was used to estimate its mass. Currently there have been several proposals for a future lepton collider, around which collaborations have formed. These include: the FCC-ee at the Swiss-French border [6], the ILC in Japan [7], and the CEPC in China [8]. This thesis focuses specifically on the FCC-ee, both in the context of the flavour tagging studies introduced below, and the subsequent characterisation of silicon detectors.

As the partons produced in a hard scattering process decrease in energy scale, they radiate a cascade of strongly interacting particles. The resulting collimated spray of colourless particles is referred to as a jet. While a jet is an interesting physics object in and of itself, offering insight into the dynamics of the strong force, such as the strength of its coupling α_s, most physics measurements ultimately target the partons produced by the hard scattering. Thus the need for a mapping between the spray of detector stable particles back to their initiating partons emerged, known as *jet-flavour tagging*. Early approaches to jet flavour tagging focused on physics-inspired, high-level observables, such as the impact parameter significance, designed to identify jets of a particular species. With the advent of sophisticated machine learning techniques, jet flavour tagging has increasingly relied on lower-level information, and high-capacity models, often exceeding 10^5 parameters, carefully trained on MC data. Among these, transformer-based architectures have recently emerged as computationally efficient models that are able to exploit subtle correlations between jet constituents.

---

[1] This is, of course, not quite true, as it neglects radiative effects such as ISR, which modify the final state. Nevertheless, the process remains remarkably clean compared to hadronic collisions.

At the FCC-ee the clean environment and well-defined initial states will provide an ideal setting for precision jet flavour tagging. In particular, the identification of strange quark initiated jets will enable direct access to processes including $\mathrm{Z} \to s\bar{s}$, $\mathrm{W} \to cs$, and rare Higgs boson decays, channels that remain experimentally challenging and imprecise. Jet flavour tagging studies at the FCC-ee relying on a transformer-based architecture are presented in Part II of this thesis.

Charged particle tracking is a central component of modern collider experiments, and arguably the most important to jet flavour tagging. As a charged particle traverses a detector its trajectory is governed by the Lorentz force. This allows for the precise determination of a particle's momentum, based on its track curvature in the presence of a B field. Moreover, measurements of energy depositions in the sensors the particle traverses allow the identification of its species via its energy loss per unit length (dE/dx). One of the most powerful applications of charged particle tracking is the reconstruction of displaced vertices due to the longer lifetimes of certain hadrons, typically dubbed secondary vertex reconstruction. Secondary vertex reconstruction plays a central role in the jet flavour tagging described above, allowing the identification of bottom or charmed hadrons in a jet, providing a direct handle on the flavour of the originating parton.

At future lepton colliders high-resolution sensors will be instrumental for precise impact parameter measurements and secondary vertex reconstruction, enabling both jet flavour tagging and a rich flavour physics programme. A primary complication as particles traverse detector material is their small-angle deflection due to multiple coulomb scattering, significantly degrading the precision of track reconstruction. Monolithic Active Pixel Sensors (MAPS) offer a compelling avenue to low-material, high-precision tracking, by combining the sensor and the readout circuitry onto the same silicon die.

While this entails a variety of advantages with respect to having dedicated sensing and readout chips, ranging from a lower material budget to miniaturisation, MAPS have seen limited adoption at hadron colliders due to their lower radiation tolerance. At lepton colliders, however, the radiation environment is foreseen to be considerably milder than at the LHC. MAPS are thus quickly emerging as *the* enabling technology for vertexing systems at future lepton colliders. As detailed later in this thesis, all four primary detector concepts proposed for the FCC-ee foresee the use of MAPS for vertexing. The characterisation of MAPS tests structures produced in the 65nm TPSCo CMOS imaging process, targeting future vertexing systems, and in particular the FCC-ee, is presented in Part III of this thesis.

This thesis begins with an overview of the general prerequisites for understanding the material herein presented. Chapter 2 starts with a brief overview of the Standard Model, followed by a discussion of some of its known limitations, which serve as the motivation for future collider projects. The details of the Future Circular Collider project are given in Chapter 3, which is primarily framed as an answer to the open questions mentioned in the previous chapter. Particular emphasis is placed on the FCC-ee's physics potential, and the detector concepts that have emerged to exploit it. Chapter 4 discusses the reconstruction of physics objects by combining different subdetector measurements, with particular emphasis on jets, serving as a segue to Part II, which details jet flavour tagging studies at the FCC-ee.

Part II begins with a pedagogical introduction to the topic of machine learning in Chapter 5, presenting the foundations before delving into topics particularly relevant for jet flavour tagging. The chapter concludes with a brief survey of machine learning in jet flavour tagging, providing context for the studies that follow. Chapter 6 subsequently presents flavour tagging studies at the FCC-ee based on fast simulation of the IDEA detector concept. Focus is placed on strange jet identification and its dependence on the identification of charged and neutral Kaons.

Part III encompasses the characterisation of MAPS test structures produced in the 65nm TPSCo CMOS imaging process. Chapter 7 presents the fundamentals of semiconductor detectors, beginning with the semiconductor band structure before delving into signal formation. The chapter ends with a brief discussion of signal readout and a presentation of MAPS principles. The details of the CE-65 family of chips are presented in Chapter 8, setting the stage for their characterisation. Chapter 9 details the lab characterisation of the CE-65v1 and CE-65v2 chips conducted at the University of Zurich, including measurements of noise, gain calibration, charge sharing, and gain uniformity. Chapter 10 subsequently details the characterisation of the CE-65v2 chips at an SPS and DESY testbeam, focusing on the achievable efficiency and spatial resolution, as well as on the ramifications of a staggered matrix geometry. The thesis concludes with Part IV, where the findings are summarised and future directions of research are highlighted.

Part I

Theoretical and Experimental Foundations

2 Standard Model

The Standard Model of particle physics represents a triumph of quantum field theory that continues to withstand experimental scrutiny. Developed throughout the 20th century, the Standard Model neatly classifies all known particles into three generations of spin 1/2 *fermions* and integer spin *bosons*. Figure 2.1 depicts the Standard Model particles and their properties. In the Standard Model picture, fermions represent the fundamental constituents of matter, while forces are represented by the exchange of vector bosons. Comprehensive treatments of the Standard Model can be found in Ref. [9]. This chapter presents a qualitative overview of its central elements and known limitations.

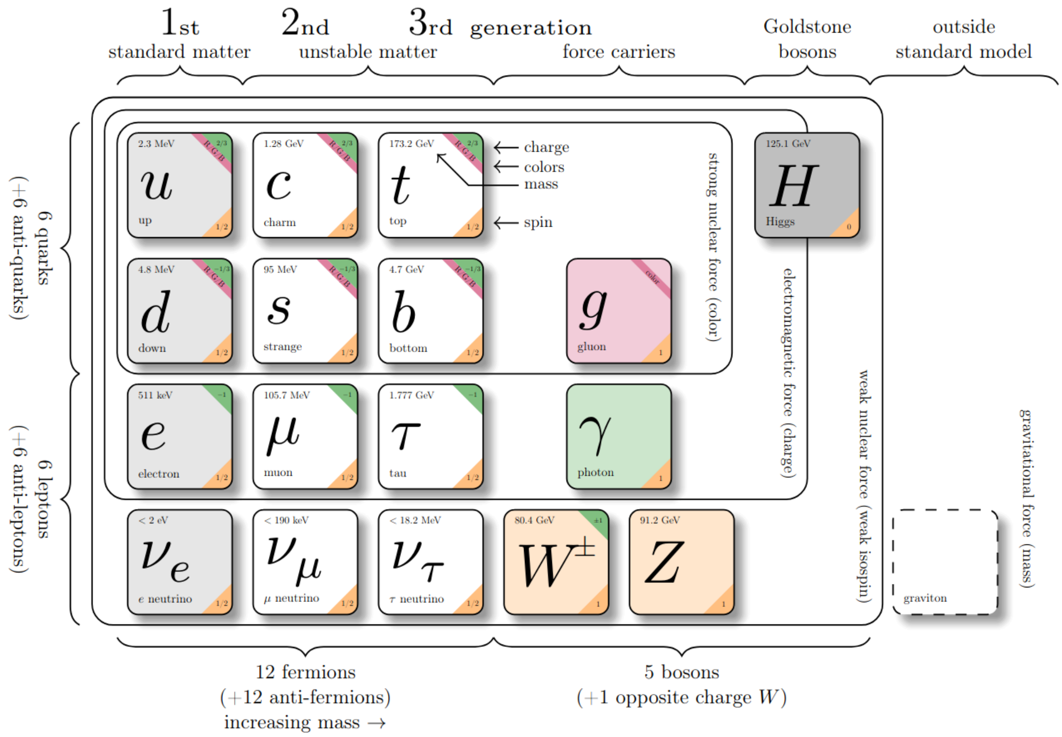

1st standard matter
2nd unstable matter
3rd generation
force carriers
Goldstone bosons
outside standard model
6 quarks (+6 anti-quarks)
6 leptons (+6 anti-leptons)
2.3 MeV u up
1.28 GeV c charm
173.2 GeV t top
charge
colors
mass
spin
125.1 GeV H Higgs
4.8 MeV d down
95 MeV s strange
4.7 GeV b bottom
g gluon
511 keV e electron
105.7 MeV μ muon
1.777 GeV τ tau
γ photon
< 2 eV ν_e e neutrino
< 190 keV ν_μ μ neutrino
< 18.2 MeV ν_τ τ neutrino
80.4 GeV W±
91.2 GeV Z
graviton
strong nuclear force (color)
electromagnetic force (charge)
weak nuclear force (weak isospin)
gravitational force (mass)
12 fermions (+12 anti-fermions) increasing mass →
5 bosons (+1 opposite charge W)

Figure 2.1: Overview of the Standard Model of particle physics. Figure taken from Ref. [10].

Matter

Fermions are the fundamental constituents of *matter*. The most familiar particle of the Standard Model is arguably the electron. It interacts electromagnetically and its low mass means it is stable. Together with its closely related, but heavier, second and third generation cousins, the muon and the tau, electrons form what are called leptons: particles which interact both electromagnetically and via the weak force. Theoretically, they are treated as part of an electroweak $SU(2)_L$ doublet together with their charge-less counterparts, the neutrinos. Neutrinos, by contrast, only interact via the weak force, and are thus notoriously difficult to detect.

Most observable matter, however, appears not as fundamental particles, but in the form of *hadrons*, or composite particles consisting of more than one quark. Hadrons appear overwhelmingly in the form of baryons, consisting of three quarks, typically of the first generation. Quarks are grouped into up-type ($+\frac{2}{3}$) and down-type ($-\frac{1}{3}$) according to their electric charge. As with leptons, each subsequent generation is considerably more massive than the one preceding it, and second and third generation quarks are orders of magnitude rarer than their first generation cousins, as their large masses render them unstable and only produced in high-energy processes. The peculiar ordering of fermion masses in the Standard Model remains an open question.

The fact that quarks are not observed as free particles, but as hadrons, is a direct consequence of the strong force, whose potential rapidly increases as quarks are separated in a process called *confinement*. Hadrons are fundamentally colour singlets, which in effect means that they do not carry any charge associated with the strong force, and are thus not subject to it. Before delving into the details of confinement, some more background is necessary, and the details are thus left to the following passages.

Forces

The vector bosons mediate three of the four known forces: the electromagnetic, the strong, and the weak force. Although the Standard Model fails to accommodate gravity, it in some sense has no bearance on collider physics, as the gravitational coupling is many orders of magnitude lower than comparable couplings of the remaining forces at 1 GeV [11]:

$$\alpha_s \sim 1 > \alpha_{QED} \approx \frac{1}{137} > \alpha_W \sim 10^{-6} \gg \alpha_g \sim 10^{-39} \tag{2.1}$$

Electromagnetism is arguably the simplest of the three Standard Model forces after symmetry breaking, as the theory that describes it, Quantum Electrodynamics (QED), is an Abelian gauge theory based on a $U(1)_{EM}$ symmetry, where the conserved quantity is the electric charge. More tangibly, this means that while the massless photon mediates electromagnetic interactions between charged particles, it does not couple to itself. This aligns with the intuitive picture of electric field lines spreading across space, where Gauss' law gives the familiar $\sim 1/r^2$ scaling extending to infinity. In the infrared limit, the coupling constant of QED is given by the fine structure constant $\alpha_{QED} = e^2/4\pi \approx 1/137$. Self-energy corrections to the photon propagator modify this value depending on the momentum transfer Q, giving rise to an increase known as the *running* of the coupling.

The strong force, by contrast, is described by Quantum Chromodynamics (QCD), which is a non-Abelian gauge theory based on an $SU(3)_C$ symmetry. While the strong force also features a massless vector boson as its mediator, the gluon, the key distinction is that the gluon carries both a charge and an anti-charge, and thus couples to itself. Moreover, rather than a single charge as in the case of electromagnetism's $U(1)_{EM}$ symmetry, the $SU(3)$ symmetry features three distinct charges, collectively denoted the colour charge, which takes the values of r, g, b. A core feature of QCD is that, in contrast to QED, the running of the strong coupling α_s *decreases* with increasing momentum transfer Q. More precisely, the running of the coupling is governed by the Beta function[1] [12]

$$\beta(g) = Q\frac{dg}{dQ} = -\frac{b_0}{(4\pi)^2}g^3 + \mathcal{O}(g^5) \tag{2.2}$$

where $g = \sqrt{4\pi\alpha}$, $b_0 = 11 - (2/3) \cdot n_f$ with $n_f = 5$ as the number of quark flavours that contribute[2], and the right-most term is the 1-loop result. The 1-loop differential equation is solved by

$$\alpha(Q) = \frac{\alpha(Q_0)}{1 + \frac{b_0}{2\pi}\,\alpha(Q_0)\ln\left(\frac{Q}{Q_0}\right)} \tag{2.3}$$

where Q_0 is some reference momentum at which the value of the coupling is known. It is the running of the strong coupling ultimately leads to jet formation, briefly described in Section 4.3. An interesting feature of the Standard Model is that the couplings seem to converge to a similar value at high energy scales, hinting at some yet-unresolved deeper symmetry.

The final force is the weak force, which is described by electroweak theory. Electroweak theory unifies electromagnetic and weak interactions based on an $SU(2)_L \times U(1)_Y$ symmetry [13–15]. The $SU(2)_L$ symmetry acts on left-handed fermion doublets and introduces three massless gauge bosons $(W_\mu^1, W_\mu^2, W_\mu^3)$, where the weak isospin is the conserved charge in interactions. The $U(1)_Y$ symmetry acts on all fermions and introduces a single massless gauge boson B_μ, with the weak hypercharge Y as the conserved charge. The breaking of the $SU(2)_L \times U(1)_Y$ symmetry by the introduction of a Higgs doublet with non-zero vacuum expectation value leads to the mixing of the gauge bosons, and the generation of mass terms for three of the four resulting gauge bosons:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \tag{2.4}$$

$$Z_\mu = \cos\theta_W\, W_\mu^3 - \sin\theta_W\, B_\mu \tag{2.5}$$

$$A_\mu = \sin\theta_W\, W_\mu^3 + \cos\theta_W\, B_\mu \tag{2.6}$$

where θ_W is the Weinberg, or weak mixing, angle, defined by the relation $\tan(\theta_W) = g_Y/g_W$ for the couplings of the $U(1)_Y$ and $SU(2)_L$ gauge groups. The A_μ is the massless photon described above, while the $W^\pm, Z$ are the resulting, physical bosons with masses of 80 GeV and 91 GeV, respectively. It is primarily the large mass that the weak gauge bosons acquire after symmetry breaking that gives the weak force its name. The virtuality of particles respects Heisenberg's uncertainty principle $\Delta E \Delta t \gtrsim \hbar$, and thus the timescale of weak interactions is very small, or in other words short range.

Perhaps the most peculiar feature of the weak force is that it is the only force that allows flavour changing decays. This results from the fact that the interaction eigenstates of charged currents do not align with the quark mass eigenstates. The change of basis matrix for quarks is known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix [16, 17] and is given by

---

[1] Technically the Beta function is defined in terms of the renormalisation scale μ, not the momentum transfer. However, a discussion on renormalisation is beyond the scope of this short introduction.

[2] The top quark is too massive to contribute at momentum transfers below hundreds of GeV.

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.7}$$

The small but non-zero off diagonal elements are in essence responsible for flavour changing weak transitions between up-type and down-type quarks of different generations. These transitions underlie the decay of heavy hadrons, and are ultimately responsible for their finite lifetimes, key in the flavour tagging discussed in Part II of this thesis.

Higgs Boson

The final element of the Standard Model is the Higgs boson, which is uniquely situated as the only scalar boson. The Higgs boson is not a force carrier like its vector boson counterparts, but rather an intimate part of the mass-generating mechanism of the Standard Model. In particular, the Higgs boson appears in the Standard Model as a complex scalar doublet

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.8}$$

The gauge symmetry of the Lagrangian is spontaneously broken by the non-zero vacuum expectation value [18, 19], with the weak vector bosons acquiring mass via the interaction terms in the covariant derivative, while the photon remains massless. Fermion masses, on the other hand, arise from the so-called Yukawa couplings: interaction terms between the Higgs field and the left- and right-handed fermion fields. After symmetry breaking these terms reduce to mass terms proportional to the Higgs vacuum expectation value

$$m_f = y_f \frac{v}{\sqrt{2}} \tag{2.9}$$

with heavier fermions, like the top quark, exhibiting the strongest Higgs couplings. The central role of the Higgs in the Standard Model, along with the unknown nature of its properties such as its potential, situate it at the very focus of current research.

2.1 Beyond the Standard Model

The Standard Model of particle physics is widely regarded as one of the most successful scientific theories ever devised. Its role in unifying three fundamental forces into a single quantum field theory, and accurately describing all known elementary particles, position the Standard Model at the very core of our understanding of the universe. It has been tested with unprecedented rigour at collider experiments, and a sizeable deviation from its predictions is yet to be observed. And yet, the present sentiment is that the Standard Model is the low-energy limit of some deeper, underlying theory.

Despite its success, the Standard Model leaves a number of observed phenomena unexplained, and does not accommodate gravity. A straightforward *quantisation of gravity* has yet to be demonstrated, due to the apparent non-renormalisability of a gravitational quantum field theory. Extensions such as String theory [20, 21] offer possible routes towards quantum gravity. Nevertheless, the path forward is unclear since the effects of gravity are negligible at the energy scales probed at current and foreseeable collider experiments. Dimensional analysis of the relevant constants reveals the Planck energy scale $E_P = \sqrt{\frac{\hbar c^5}{G}} \sim 10^{19}$ GeV at which quantum corrections to classical gravity are expected to become relevant, which is many order of magnitude higher than even the reach of the LHC. The unification of gravity and the Standard Model thus remains one of the primary open questions of modern physics.

A number of cosmological measurements appear to be incompatible with the Standard Model and general relativity. The rotation curves of spiral galaxies [22, 23], for instance, appear to deviate from expectations based on the luminous mass, which is located primarily in the centre, displaying instead a linear $M(r \leq R) \sim R$ dependence, hinting at some form of invisible matter that does not interact electromagnetically, dubbed *dark matter*. Roughly 85% of matter in the universe is thought to be dark matter, and yet the Standard Model offers no compelling dark matter candidates. A class of models extending the Standard Model to include weakly interacting massive particles (WIMPs) postulates dark matter candidates whose small cross sections and subtle signatures have meant they have thus far remained undetected. Experiments at the LHC are at the forefront of WIMP searches, revolving around experimental signatures such as missing energy, as well as deviations in precision measurements.

A well-motivated extension of the Standard Model, supersymmetry, introduces an additional symmetry between fermions and boson, where every Standard Model particle has an associated superpartner whose spin differs by 1/2. While supersymmetry introduces dark matter candidates addressing the problem stated above, the primary motivation lies in resolving the problem of *naturalness*, whereby the Standard Model requires the precise cancellation of the bare Higgs mass and its large self-energy corrections to 1 part in 10^{34} in order to obtain the comparatively miniscule pole mass of $m_H \approx 125$ GeV. Supersymmetry instead addresses the observed Higgs mass by introducing cancellations between loop corrections of Standard Model particles and their superpartners. Despite its motivations, the LHC has placed stringent limits on supersymmetry, excluding many models and placing lower bound on superpartner masses.

Beyond gravity, dark matter, and naturalness, further glaring limitations of the Standard Model include the observed non-zero neutrino masses [24–26] and the matter-antimatter asymmetry of the universe. However, the Standard Model does not accommodate massive neutrinos, since the absence of right-handed neutrinos forbids a Dirac mass term. Extensions including treating the neutrino as its own antiparticle exist, but minimally require a modification of the Standard Model. While the Standard Model does accommodate some matter-antimatter asymmetry via CP violation, it does not suffice for the observed baryon asymmetry. The precise properties of the Higgs boson, including the shape of its potential, may hold clues regarding the prerequisite conditions for baryogenesis.

Thus the particle community is rallying behind a number of proposed "Higgs factories", or lepton colliders where Higgs bosons may be plentifully produced in a clean experimental environment. These include linear collider concepts, such as the ILC [7] and CLIC [27], which generally feature high centre-of-mass energies, albeit with reduced luminosity, and circular colliders, such as CEPC [8] and the FCC-ee [6]. The following chapter introduces the FCC project in detail, outlining its design, physics goals, and proposed detector concepts.

3 The Future Circular Collider Project

Colliders have historically served as engines of discovery, providing a way forward in tandem with theoretical efforts, and hinting at the existence of yet-undiscovered phenomena. Today however, the field of particle physics is at a crossroads. Following decades of success, the field has shifted from targeted discoveries to precision tests of the Standard Model, and studying any deviations therefrom. The discovery of the Higgs boson at the LHC marks its most recent triumph, but does not chart a clear course forward. The Future Circular Collider (FCC) project [28, 29] has emerged as a two-pronged answer to the open questions the field currently faces, tackling present challenges both at the *precision* and *energy* frontiers.

This chapter begins with a brief introduction into the fundamentals of collider physics, before presenting the Future Circular Collider in Section 3.2. The remainder of this chapter details several aspects of the FCC-ee, including the proposed runs in Section 3.3 and physics programme in Section 3.4, before concluding by presenting the detector concepts that have emerged to address these challenges.

3.1 Introduction to Collider Physics

Particle colliders function by accelerating particles to high energies and colliding them in bunches. The resulting interactions produce on-shell particles that manifest as final state objects which can be directly studied. Additionally, the scattering process is influenced by off-shell virtual particles, offering an avenue to indirectly infer their properties, often at higher energy scales. Particle colliders can be broadly divided into lepton colliders, which collide electrons and positrons[1], and hadron colliders, which collide protons and (anti-)protons. Additional collider types featuring linear geometries or alternative particle combinations (e.g. eh, heavy ion, ...) exist and have played significant roles in advancing the field, particularly in contexts like deep inelastic scattering or QCD matter at high temperatures and densities, but will not be discussed for brevity. Presently, the largest particle collider is the LHC at CERN, which collides hadrons at energies of up to 13.6 TeV.

Hadron and lepton colliders differ greatly in their motivation. While the more massive hadrons allow collisions at considerably higher energies, the collision environment is substantially more challenging as hadrons are not fundamental, and their constituents carry colour charge. Lepton colliders, on the other hand, are fundamentally precision machines.

The precise energy at which a particle collider operates is defined with respect to the Mandelstam variable s, given as

$$s = (p_1^\mu + p_2^\mu)^2 = (p_1^\mu + p_2^\mu)(p_{1\mu} + p_{2\mu}) = (E_1 + E_2)^2 - |\vec{p}_1 + \vec{p}_2|^2 \tag{3.1}$$

where $p_{1,2}^\mu$ are the 4-momenta of the incoming particles in a 2-body scattering process. In the centre-of-mass frame, where the total 3-momentum vanishes, this reduces to $\sqrt{s} = E_1^{\mathrm{CM}} + E_2^{\mathrm{CM}}$, and defines the centre-of-mass energy of the collision, interpreted as the total energy available for particle production. Operating at higher centre-of-mass energies broadens the collider's physics reach, as it allows the on-shell production of more massive particles and access to new kinematic regimes. However, increased energy entails notable drawbacks, including the need for stronger bending magnets, increased synchrotron radiation, and substantially higher energy consumption.

---

[1] Ref. [30] provides a review of recent efforts exploring the feasibility of a multi-TeV muon collider.

Beyond the centre-of-mass energy, the second key parameter is the number of events that is produced for the physics process under consideration. The number of events depends both on the cross section σ, which characterises the probability of a process, and on the luminosity $\mathcal{L}$. The *luminosity* is a measure of the number of particles through a unit area per unit time, quantifying the intensity of beam overlap. It is given in units of inverse area (e.g. cm^{-2}) and inverse time (e.g. s^{-1}). In the case of the head-on collision of two Gaussian beams, the luminosity is given by [31]

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \tag{3.2}$$

where $N_{1,2}$ are the number of particles per bunch for the colliding beams, N_b is the number of bunches per beam, f is the revolution frequency, and $\sigma_{x,y}$ are the transverse beam sizes. The number of events can be obtained by integrating the product of the luminosity and the cross section, given as

$$N_{ev} = \int \sigma \cdot \mathcal{L}\, dt = \sigma \cdot \mathcal{L}_{\text{int}} \tag{3.3}$$

where N_{ev} is the number of produced events for the given process and $\mathcal{L}_{\text{int}}$ is the integrated luminosity. With the basic concepts of centre-of-mass energy and luminosity established, the following section discusses how they inform the design of future facilities such as the FCC.

3.2 The FCC Integrated Project

The Future Circular Collider (FCC) project is an ambitious two-stage project that envisions the construction of a 90.7 km tunnel at CERN, to successively host a high luminosity electron-positron collider, the FCC-ee, followed by a high energy hadron-hadron collider, the FCC-hh. It will serve as the successor to the LHC, and define the future of particle physics for the latter half of the 21st century. Figure 3.1 depicts the proposed geometry for the FCC-ee accelerator. Additional civil engineering constraints to host both colliders in the same tunnel are being considered in the design phase, including an increase in the tunnel diameter, with respect to the LHC, to 5.5 m.

The natural synergy observed between LEP and the LHC serves as a primary motivation for the two-stage approach adopted for the FCC project. As a lepton collider, LEP operated in a comparatively clean environment, instrumental in the precise determination of a variety of Standard Model parameters, particularly related to the Z boson, such as the mass (m_Z) and width (Γ_Z), which remain state-of-the-art. Moreover, and just as importantly, is the oft-touted phrase that "precision is discovery". High-mass particles enter lower energy precision measurements though virtual loop corrections, shifting observables away from their expected values in the absence of such contributions, as exemplified in Section 3.4. Thus, the precise measurement of Standard Model parameters, and potential deviations from theoretical predictions, can provide sensitivity to heavier states, helping chart the course forward for future high-energy hadron

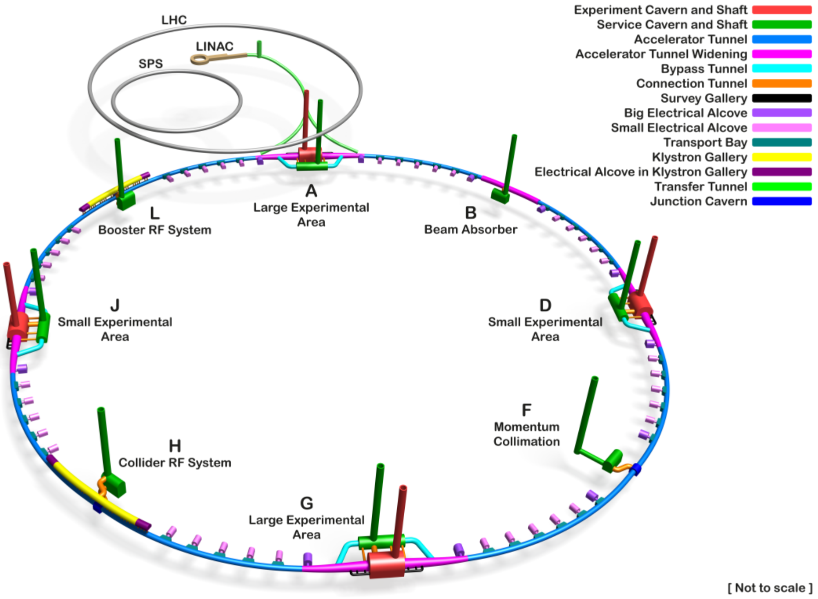

LHC
LINAC
SPS
Experiment Cavern and Shaft
Service Cavern and Shaft
Accelerator Tunnel
Accelerator Tunnel Widening
Bypass Tunnel
Connection Tunnel
Survey Gallery
Big Electrical Alcove
Small Electrical Alcove
Transport Bay
Klystron Gallery
Electrical Alcove in Klystron Gallery
Transfer Tunnel
Junction Cavern
A
Large Experimental Area
B
Beam Absorber
L
Booster RF System
J
Small Experimental Area
D
Small Experimental Area
F
Momentum Collimation
H
Collider RF System
G
Large Experimental Area
[Not to scale]

Figure 3.1: Schematic of the FCC-ee accelerator. The four collision points are marked in red. Figure taken from Ref. [28].

colliders.

The FCC-hh is planned to begin operations in the 2070's with proton-proton collisions at a baseline energy of 85 TeV, potentially increasing to $\sim$ 120 TeV, depending on ongoing R&D. Operating at nearly an order of magnitude higher centre-of-mass energy than the LHC, the FCC-hh will enable direct searches for heavy resonances at energy regimes inaccessible to its predecessor. Additionally, the copious Higgs boson production, expected to exceed 20 billion [29], will constrain rare Higgs decays including $H \to \gamma\gamma, \mu\mu, Z\gamma$ at unprecedented precision, and facilitate the precise measurement of the Higgs self-coupling via double-Higgs production. Many of the proposed detector and accelerator technologies for the FCC-hh require significant R&D, a challenge made tractable by its distant timeline. One of the most critical areas for development are the high-field superconducting magnets required for the 100 TeV FCC-hh collisions [29].

In comparison, the preceding stage, the FCC-ee, relies on more mature and well-established technologies. The FCC-ee is optimised for precision physics at lower centre-of-mass energies, ranging from the Z-pole ($\sim$ 90 GeV) to just above the $\mathrm{t\bar{t}}$ production threshold ($\sim$ 365 GeV) during its 15 year runtime, presumably beginning around 2045. Its rich physics programme spans Higgs physics, precision electroweak measurements, flavour physics, precision QCD, and even light BSM searches, as detailed in Section 3.4. Moreover, the FCC-ee features the highest number of collisions of the proposed colliders at the production thresholds of the Z boson, W boson pairs, and a Higgs in association with a Z. Thus it offers the most exquisite precision, and has garnered much support in Europe and beyond.

3.3 FCC-ee Parameters

The FCC-ee will be hosted in the 90.7 km tunnel described above, and will feature four interaction points delivering unprecedented luminosities across a range of energies. The collider will operate at a variety of different beam configurations depending on the targeted centre-of-mass energy, requiring the tuning of a number of parameters including beam current, bunch multiplicity and spacing, as well as transverse bunch dimensions. During the Z-pole run the FCC-ee is expected to achieve a luminosity of $145 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ per interaction point, delivered as 11200 bunches per beam with a spacing of 27 ns. Detailed beam parameters for other energies can be found in Ref. [29].

A summary of the proposed run schedule detailing the centre-of-mass energies, integrated luminosity, and associated physics targets is depicted on Figure 3.2. It should be noted that Figure 3.2 assumes the baseline scenario, with the precise order and duration not yet fixed [29]. The FCC-ee will explore the vicinity of the Z-pole during its initial four year run at centre-of-mass energies of $\sqrt{s} = 88, 91, 94$ GeV, during which an integrated luminosity of 205 ab^{-1} will be achieved. This corresponds to a production of 6×10^{12} Z bosons, offering unprecedented statistics for the measurement of Z boson properties and rare flavour physics decays. Following the Z pole run, the FCC-ee will operate at centre-of-mass energies corresponding to the WW-threshold at $\sqrt{s} = 157, 163$ GeV for two years. The integrated luminosity of 19.2 ab^{-1} corresponds to the production of 2.4×10^8 W boson pairs, a factor $\sim 5 \times 10^3$ higher than at LEP, allowing the precise study of W boson properties, improved determinations of CKM elements, and detailed studies of non-perturbative QCD effects in WW decays, detailed in the following section.

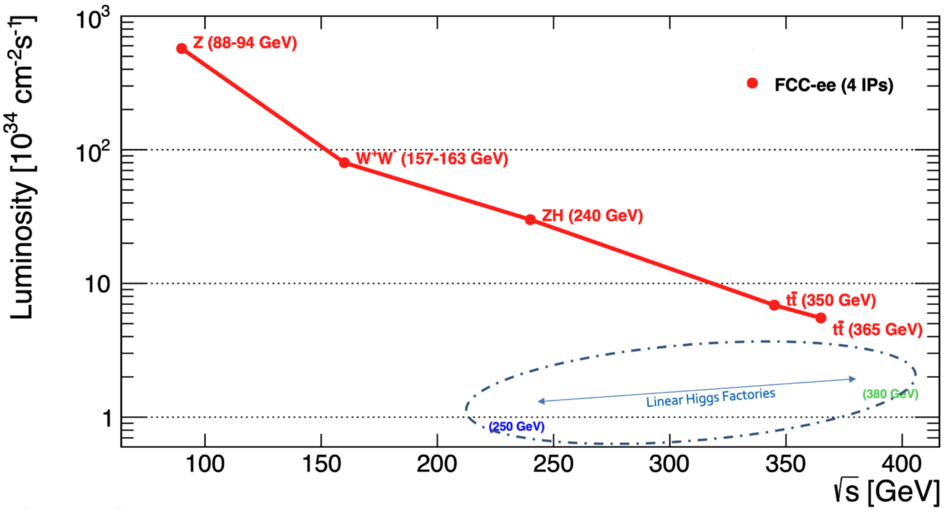


Figure 3.2: The integrated luminosities to be delivered in the baseline run schedule of the FCC-ee, shown as a function of centre-of-mass energy. A luminosity range for linear colliders is shown above 200 GeV. Figure taken from Ref. [29].

The Higgs programme of the FCC-ee begins with the ZH run at $\sqrt{s} = 240$ GeV. The three year run will result in an integrated luminosity of 10.8 ab^{-1}, corresponding to 2.2×10^6 Higgsstrahlung and 6.5×10^4 WW fusion events.

The $\mathrm{t\bar{t}}$ run is sub-divided into a 1 year run scanning the production threshold at $\sqrt{s} = 340 - 350$ GeV, and a 4 year run slightly above the threshold at $\sqrt{s} = 365$ GeV. The threshold scan will deliver an integrated luminosity of 0.42 ab^{-1}, allowing a precise determination of the top quark mass (m_t) and decay width (Γ_t). The 365 GeV will deliver an integrated luminosity of 2.70 ab^{-1}, with the primary goal of measuring top couplings, and offering indirect sensitivity to the Higgs self coupling when combined with the ZH run. Across both $\mathrm{t\bar{t}}$ runs, a total of 2×10^6 $\mathrm{t\bar{t}}$ pairs, 3.7×10^5 Higgsstrahlung events, and 9.2×10^4 WW fusion events will be produced.

3.4 Physics Programme

The full physics programme of the FCC-ee is much too detailed to be explained in the space herein alloted, so the interested reader is encouraged to consult Refs. [6, 29] for a more complete discussion. As mentioned in Section 3.2, the FCC-ee physics programme spans Higgs physics, precision electroweak measurements, flavour physics, precision QCD, and even light BSM searches. Some aspects of these will be highlighted.

The *Higgs* programme at the FCC-ee is centered around the measurement of the Higgs boson mass, constraints on its total width, as well as the precise determination of many of its couplings to Standard Model particles. The recoil mass method exploits the well-determined kinematics of a lepton collider to reconstruct the invariant mass of the system recoiling against the $Z \to l^+l^-$ produced in Higgsstrahlung events, allowing the identification of Higgs events without reference to a specific decay mode. The peak position of the recoil mass will allow the determination of the Higgs mass (m_H) to high precision. The recoil mass method additionally provides a measurement of the total ZH cross section σ_{ZH}, which when coupled with the $H \to ZZ^*$ branching fraction allows the determination of the Higgs width Γ_H [29]. The FCC-ee will improve the precision on a number of κ parameters, which quantify deviations of Higgs couplings from their Standard Model predictions, with respect to HL-LHC projections, with κ_Z improving by over an order of magnitude, and more moderate gains expected for couplings to the W, b, c, g, τ, μ, γ, $Z\gamma$ [29]. Beyond this, the prospect of constraining the electron Yukawa coupling [32] with a dedicated run at the Higgs boson pole mass is being actively considered. While not explicitly included in the baseline run plan depicted on Figure 3.2, a dedicated $e^+e^- \to H$ run would allow a $\sim 1.6\times$ Standard Model value upper limit to be set within 2 years [32]. The s-channel Higgs production presents one of the most formidable challenges to the FCC-ee due to the narrow Higgs boson width. An excellent control over the beam energy spread would be required, lest the Higgs resonance be rendered invisible.

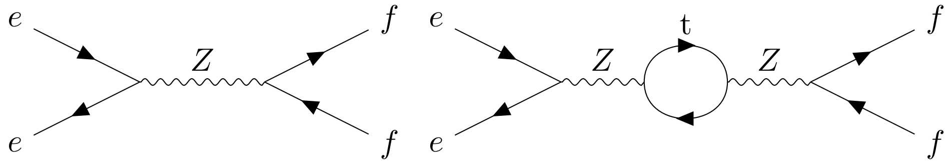
The *precision electroweak* programme at the FCC-ee is complemented by the Z-pole, WW threshold, and $\mathrm{t\bar{t}}$ run. In particular, the exquisite statistics in the vicinity of the Z-pole will allow a precise determination of both the Z boson mass (m_Z) and width (Γ_Z) from the line shape scan, improving present determinations by over 1 and 2 orders of magnitude, respectively [29]. A multitude of additional precision observables will become accessible during the tera-Z run, including the weak mixing angle and the b-quark forward backward asymmetry (A^b_{FB}), likewise improving by over two orders of magnitude. Improvements of a similar scale are expected the W mass (m_W) and width (Γ_W) from the WW threshold scan. Rapid advances in flavour tagging precision, such as those presented in Chapter 6, will additionally allow the direct extraction of the $|V_{cs}|$ and $|V_{cb}|$ CKM matrix elements from hadronic W decays, as was done at LEP in the case of the former [33–35]. The $\mathrm{t\bar{t}}$ threshold scan will improve present determinations of the top quark mass (m_{t}) by almost two orders of magnitude, instrumental for a precise determination of a number of electroweak precision observables, where it enters as a parametric uncertainty [29]. The top quark width (Γ_{t}) will likewise be extracted from the $\mathrm{t\bar{t}}$ threshold scan, albeit with milder improvements in precision. The high precision measurement of the top quark's properties will substantially tighten the global electroweak fit, improving constraints on Standard Model parameters and exposing any potential tensions that may arise.

The FCC-ee is seldom framed as a QCD machine, but offers a clean environment and exquisite statistics, accommodating a rich *precision QCD* programme. The collision of colourless particles fixes the initial state kinematics, and allows quarks and gluons to hadronise in a relative

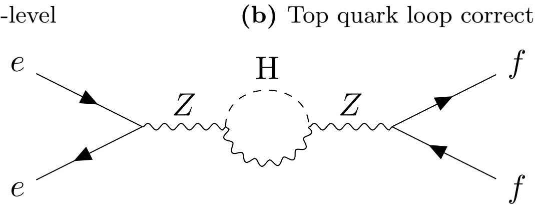
"QCD vacuum", devoid of partons unrelated to the main physics process. The determination of the strong coupling is paramount to improving the precision of virtually all cross sections and decays at colliders, where it can enter either directly (e.g. $gg \to \mathrm{t\bar{t}}H$), or via higher order corrections (e.g. $e^+e^- \to Z(q\bar{q})$). At the FCC-ee, the strong coupling α_s is expected to be the leading parametric uncertainty in the determination of the Z boson width described above [36]. Nevertheless, the FCC-ee is uniquely suited for a per-mille level determination of the strong coupling at the Z-pole ($\alpha_s(m_Z)$) through combined fits of the hadronic-to-leptonic decay ratio ($R_Z(Q)$), the total width (Γ_Z), and the hadronic cross section ($\sigma_Z^{\text{had.}}$). The strong coupling could similarly be determined at the W boson pole ($\alpha_s(m_W)$) from hadronic W decays, offering an improvement of around two orders of magnitude with respect to present determinations [29]. Additional determinations of the strong coupling are possible via hadronic τ decays, as well as event shapes and jet rates [37, 38], as was done at LEP for the latter [39]. The FCC-ee also offers a plethora of opportunities to study non-pQCD effects, such as colour reconnection in WW events, or gluon radiation patterns from an ultra-pure gluon jet sample ($ZH(\to gg)$).

The *flavour physics* programme at the FCC-ee is centered around the exquisite statistics at the Z-pole, and will cover a wide variety of topics ranging from searches for flavour-changing neutral currents to the precision study of bottom and charmed hadrons. Moreover, the tera-Z run will produce over 10^{11} boosted τ leptons, corresponding to an improvement of over an order of magnitude in lepton flavour universality tests [29].

The potential of the FCC-ee as *discovery* machine is often understated. Besides probing extensions of the Standard Model that feature direct signatures, such as long-lived particles, the improvements by orders of magnitude in a number of precision observables will offer a hint at physics beyond the Standard Model. Deviations from values predicted by the Standard Model would offer compelling evidence of mass states that are yet to be identified. The signature of yet-undiscovered particles in precision measurements at energies substantially lower than their pole mass has significant historical precedent. The mass of both the top quark and the Higgs boson were predicted prior to their discovery through their influence on the cross section of the $e^+e^- \to f\bar{f}$ process via loop corrections, as depicted on Figure 3.3. In similar fashion, the FCC-ee is expected to offer sensitivity to particles with masses up to 10–70 TeV [6].



(a) Tree-level **(b)** Top quark loop correction



(c) Higgs loop correction

Figure 3.3: Tree-level (a) and 1-loop corrections involving the top quark (b) and the Higgs boson (c) to the $e^+e^- \to Z \to f\bar{f}$ scattering amplitude. Figure inspired by Figure 1.3 of Ref. [6].

3.5 Detector Concepts

The rich physics programme at the FCC-ee imposes a stringent set of performance requirements on the rapidly evolving detector concepts. For instance, beam collimation considerations at the Z-pole run limit the magnetic field strength to 2 T, requiring large tracking detector volumes to compensate [29]. The detector requirements of the FCC-ee are detailed in Refs. [29, 40, 41]. The present discussion focuses on a selection of the more optimistic targets outlined in Ref. [29], along with their motivation. Table 3.1 lists the key performance requirements on each sub-detector in the FCC-ee detector concepts.

Subsystem	Requirement	Subsystem	Requirement
Vertex	$\sigma(d_0) = 3 \oplus \frac{15}{p\,\sin^{3/2}\theta}\,\mu\text{m}$ $X/X_0 < 1\%$	**HCAL**	$\frac{\sigma_E}{E} = \frac{30\%}{\sqrt{E}}$, $\Delta x \times \Delta y = 2 \times 2\,\text{mm}^2$
Tracking	$\sigma_p/p < 0.1\%$ for $O(50)$ GeV	**Muons**	μ ID for $p < 1$ GeV
ECAL	$\frac{\sigma_E}{E} = \frac{3\%}{\sqrt{E}}$, $\Delta x \times \Delta y = 2 \times 2\,\text{mm}^2$	**Particle ID**	3σ K/π separation up to $p < 40\,\text{GeV}$

Table 3.1: Selected detector requirements for the FCC-ee detector concepts. Values adapted from "Aggressive" requirements given in Ref. [29].

The requirements on the *vertex detector* are centered around the transverse impact parameter resolution and the minimisation of the material budget in order to suppress multiple coulomb scattering, as described in Section 7.2.1. A precise reconstruction of secondary vertices is instrumental not only to heavy-flavour jet tagging, where the presence of secondary vertices helps to discriminate b-, c-, and light jets, but also to flavour physics, with the reconstruction of the thus-unseen $B \to K^*\tau\tau$ decay placing some of the most stringent requirements on the transverse impact parameter resolution. In particular, the impact parameter resolution itself consists of an asymptotic term, assumed to be 3 µm in the optimistic case, which can be improved by moving the first vertex layer as close to the beampipe as possible, and using small-pitch pixel sensors in order to achieve the exquisite spatial resolution required per layer. Moreover, the ambitious target of maintaining the entire material budget of the vertex detector below 1% X/X_0 is instrumental to minimising multiple scattering. Recent developments in Monolithic Active Pixel Sensors (MAPS) have demonstrated the feasibility of combining the twin goals of high-resolution and ultra-low material budget in vertexing systems. Some developments targeting vertexing at future colliders are detailed in the latter part of this thesis. All of the detector concepts presented in this Section envision the use of MAPS for vertexing.

The requirement on the *tracking detector* is expressed in terms of the momentum resolution for energetic particles, and is driven primarily by the measurements of the Higgs recoil mass (m_H^{recoil}), the Z boson decay width (Γ_Z), and searches for lepton flavour violating decays ($Z \to \tau\mu$), all of which require exquisite momentum measurements of energetic leptons. Both the *electromagnetic calorimeter (ECAL)* and the *hadron calorimeter* envision a high transverse resolution, expressed as the granularity $\Delta x \times \Delta y$, with the requirement being primarily driven by the reconstruction of $\pi^0 \to \gamma\gamma$ decays for the electromagnetic calorimeters, and the precise reconstruction of neutral jet constituents for the hadron calorimeter. The additional energy resolution requirement in electromagnetic calorimeter is driven by the need to accurately reconstruct photon final states, such as $e^+e^- \to \nu_e\bar{\nu}_e\gamma$, while in the hadron calorimeter it is driven by the jet energy resolution, particularly for $H \to$ 2-jet decays. The primary performance metric for the *muon system* is the muon identification performance, as precise momentum measurements are already provided by the tracking detector. *Particle identification (PID)* has additionally emerged as a performance requirement. As evidenced by the jet flavour tagging studies reported in Chapter 6 and other prior studies, the discrimination of $K^\pm/\pi^\pm$ is instrumental in the identification of strange quark initiated jets, and the promising novel physics studies thereby enabled (e.g. $H \to s\bar{s}$).

From the above requirements four primary detector concepts have emerged either as adaptions of detectors designed for other proposed lepton colliders, or as dedicated developments for the FCC-ee. With four interaction points, the FCC-ee will accomodate a wide variety of detector technologies and concepts, optimising slightly different aspects of particle detection. The IDEA and the CLD detector concepts are the most studied among the detector concepts under consideration, and are thus listed in most detail.

3.5.1 IDEA

The Innovative Detector for e$^+$e$^-$ Accelerator (IDEA) [42–44] detector concept emerged as a standalone development for the FCC-ee, combining novel subdetector designs. Figure 3.4 depicts a schematic cross section of the upper-right quadrant of the IDEA detector. While this iteration of the IDEA detector will be detailed in this section, it should be noted that the detector concepts are rapidly evolving and being optimised, thus the precise positions and details of individual subdetectors are subject to change. Beginning at the beampipe and moving radially outwards, the IDEA detector concept consists of a silicon vertex detector, a low-mass drift chamber surrounded by a silicon tracking layer, a crystal dual-readout electromagnetic calorimeter, a superconducting magnet, a dual-readout hadronic calorimeter, and finally a muon system [29].

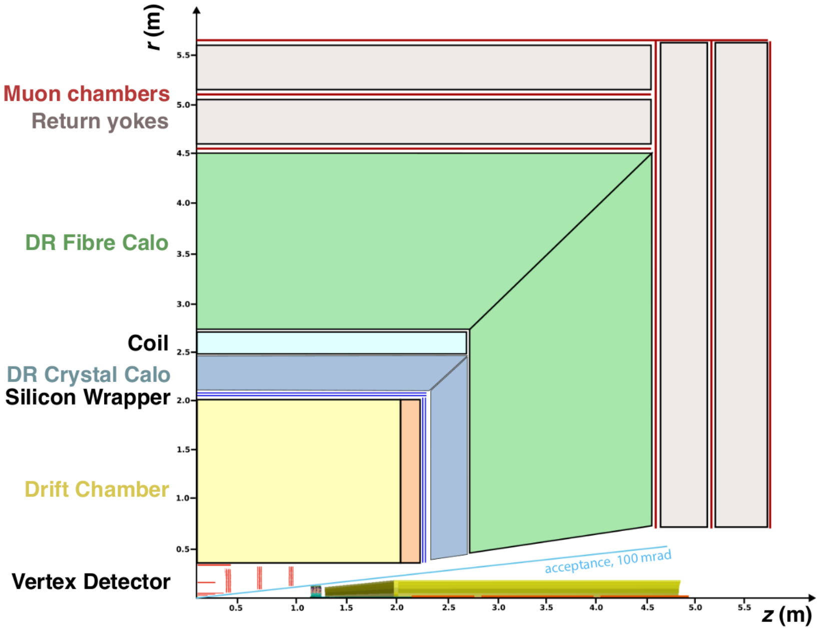


Figure 3.4: Schematic cross section of the upper-right quadrant of the IDEA detector concept. Adapted from Ref. [29].

The silicon *vertex detector* barrel consists of 5 layers of sensors extending from $r = 1.37$ cm to $r = 31.5$ cm, where r denotes the radial distance from the beampipe. The vertex detector can be further sub-divided into the inner and outer subsystems, where the inner subsystem comprises the three barrel layers closest to the beampipe at radii between $r = 13.7$ mm and $r = 35.6$ mm, respectively, while the outer subsystem comprises the outer two layers at radii of $r = 130$ mm and $r = 315$ mm and the three endcap layers at each detector end. The baseline design of the inner subsystems assume 25 µm × 25 µm MAPS based on the current ARCADIA prototype [45]. Each layer of the inner subsystem contributes 0.25% X_0 radially. An alternative ultra-light design mimicking the ALICE ITS3 upgrade [46] features fully-stitched MAPS sensors in a self-supporting structure, allowing a reduction of ~ 3 in the material budget per layer to 0.07% X_0 [47]. The outer subsystem assumes hybrid pixel sensors featuring a pixel pitch of 50 µm × 150 µm and a similar sensor thickness to the inner subsystem MAPS of ~ 50 µm.

The second part of the IDEA tracking system is the low-mass *drift chamber*. It consists of 112 layers of wires spanning 4 m along the z axis at radii between $r = 0.35$ m and $r = 2$ m in a 90% He and 10% iC_4H_{10} mixture. The drift chamber offers a low-material budget alternative to full-silicon tracking detectors, with a budget of only 1.6% X_0 for the entire subdetector, orthogonally to the beampipe. Besides providing high-precision tracking measurements, the drift chamber is expected to provide exquisite particle identification capabilities for a wide momentum range using the cluster counting technique (dN/dx). The drift chamber is surrounded by a large area silicon tracking system, providing an additional high-precision spatial measurement, and potentially a time-of-arrival measurement.

The final subdetector lying within the superconducting coil 2 T magnet is the crystal *electromagnetic calorimeter* providing energy and position measurements for electrons and photons. The electromagnetic calorimeter extends from $r \approx 2$ m to $r = 2.5$ m

The penultimate subdetector is the *hadronic dual-readout calorimeter* spanning from $r = 2.7$ m to $r \approx 4.5$ m. The dual-readout calorimeter consists of alternating rows of scintillating and Cherenkov fibres in capillary tubes. An excellent energy resolution of $30\%/\sqrt{E}$ is expected for hadrons. While the subdetector effectively functions as the hadron calorimeter in the presented design of the IDEA detector concept, its use as a combined electromagnetic and hadron calorimeter, distinguishing between the two via the ratio of scintillation and Cherenkov light, was originally the baseline and is still being actively explored.

The final subdetector is the *muon system* consisting of layers of chambers embedded in the iron return yoke surrounding the solenoid. The baseline technology is derived from the μ-RWELL muon detector system [48]. While the muon system is primarily designed for muon detection, as its name suggests, its potential for detecting LLPs is currently being explored.

It is worth highlighting two key differences between the IDEA detector concept presented in this section and the configuration used for the flavour tagging studies in Part 6:

- First vertex detector layer was placed at a radius of 1.7 cm, rather than the current, reduced baseline of 1.37 cm
- Absence of a dedicated dual-readout electromagnetic calorimeter

3.5.2 CLD

The CLIC-like detector (CLD) [49] is an adaptation of the most recent CLIC detector model to the FCC-ee. A number of modifications in order to accomodate the FCC-ee conditions was necessary, including an increase in the tracking detector volume in order to accommodate the weaker 2 T field strength. Figure 3.5 depicts a schematic cross section of a transverse slice and the upper-right quadrant of the CLD detector concept. The CLD detector features a full-silicon tracking detector, surrounded by an electromagnetic calorimeter and a hadron calorimeter, both lying within the superconducting solenoid. Surrounding the coil is the muon system.

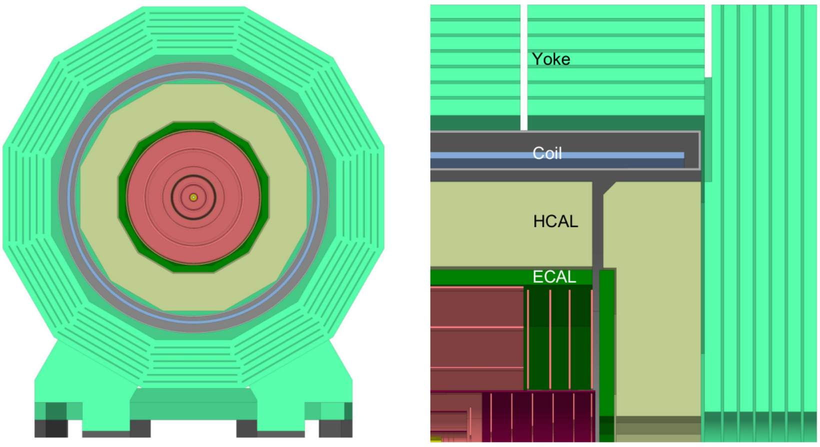


Figure 3.5: Schematic cross section of a transverse slice (a) and the upper-right quadrant (b) of the CLD detector concept. Adapted from Ref. [29].

The silicon pixel *vertex detector* of the CLD detector concept is organised in double-layers of ultralight MAPS sensors. The barrel consists of three double-layers located at $r = 12.5$-13.5 mm, $r = 37$-38 mm, and $r = 57$-58 mm, respectively. Each barrel double-layer contributes a material budget of 0.63% X_0. The three endcap double-layers are located at $z = 160$ mm, $z = 230$ mm, and $z = 300$ mm, where z denotes the longitudinal distance from the interaction point. The endcap double-layers contribute a material budget of 0.70% X_0 each.

The *tracking detector* surrounding the vertex detector is likewise comprised of silicon detectors, and is sub-divided into the inner tracker and outer tracker. The inner tracker is comprised of three barrel layers and seven endcap layers, while the outer tracker adds three additional barrel layers and four endcap disks. Each tracker layer contributes a material budget of $\sim 1.2\%$ X_0.

It should be noted that at the time of writing the inclusion of a novel ring-imaging Cherenkov detector surrounding the tracking detector is being considered. The detector would supplement the particle identification capabilities of the CLD detector, but requires a reduction in the tracking detector radius, deteriorating the momentum resolution [50].

The *electromagnetic calorimeter* and the *hadron calorimeter* both lie within the superconducting solenoid, enabling a precise reconstruction of particle trajectories in the particle-flow approach. The electromagnetic calorimeter employs tungsten as the absorber material, with silicon sensors providing a high-granularity readout. The hadron calorimeter utilises alternating steel and scintillator tiles.

The final subdetector, located outside the coil, is the *muon system* comprised of seven resistive plate muon chambers.

3.5.3 ALLEGRO

The ALLEGRO [51] detector concept, short for A Lepton-Lepton collider Experiment with Granular Read-Out, is the most recent of the presented detector concepts. Its main innovation is a high-granularity electromagnetic calorimeter using a noble liquid as the active material. The interested reader is referred to Ref. [29] for a detailed schematic of the detector layout.

The *vertex detector* is expected to make extensive use of the promising MAPS developments, as for the detector concepts already presented. The baseline vertex detector design is that of the IDEA detector concept, and will thus not be repeated. Surrounding the vertex detector is the *tracking detector*, whose precise technology is still being explored. Fundamentally, two avenues are being explored: a full-silicon tracker, as proposed for CLD, and a gaseous tracker either in the form of a straw tracker or a drift chamber, as proposed for IDEA, with the latter serving as the baseline. The *electromagnetic calorimeter* is situated within the superconducting coil, and will consist of a combination of a noble liquid as an active medium and a metallic absorber as the passive material. Proposed combinations include tungsten or lead as absorbers paired with liquid argon, or tungsten with liquid krypton, with lead and liquid argon currently serving as the baseline. The *hadron calorimeter* situated around the coil will adopt a design combining steel absorber plates and scintillator tiles, much like CLD. Finally, technology options for the *muon system* are still under active investigation, with drift tubes, resistive plate chambers, and micromegas among those proposed.

3.5.4 ILD

The International Large Detector (ILD) [52] is a detector concept originally developed for the ILC, but more recently adapted for the FCC-ee. As the CLD detector concept was developed from the ILD [29] it shares significant overlap in design choices. In particular, both detector concepts feature high-granularity electromagnetic and hadronic calorimeters within the superconducting solenoid, and envision the use of MAPS for their high-precision vertexing detectors. The primary distinction between the two detector concepts is the adoption of a time-projection chamber as the tracking detector for the ILD. While a time-projection chamber features a very low material budget and allows excellent charged particle identification via ionisation clusters, the concept still faces significant challenges for FCC-ee operation [29], particularly at the Z-pole.

Having surveyed the principal detector concepts under consideration for the FCC-ee, the following chapter provides an overview of how measurements of subdetectors are combined into reconstructed physics objects.

4 Reconstruction

The raw output of detectors, such as the concepts presented in Section 3.5, is a large number of localised signals resulting from particle interactions with different subdetectors. However, it is not until these measurements are combined and interpreted as reconstructed particles, with an associated charge and 4-momentum, that meaningful physics can be extracted. This chapter provides a very brief overview of the reconstruction chain, beginning with individual hits in the tracking system and culminating in high-level physics objects like jets.

4.1 Tracks

Charged particle tracking forms the backbone of event reconstruction at collider experiments. The sequential measurement of charged particles as they traverse finely-segmented detectors allows the reconstruction of their trajectories. The trajectory of a charged particle is bent into a helix by the Lorentz force, with a curvature that is inversely proportional to the particle's momentum. The determination of a particle's trajectory relies at minimum on two[1] points which can be interpolated. It can be shown that for an idealised two layer tracking detector the impact parameter resolution σ_b is approximately given by [53]

$$\sigma_b = \left(\frac{r_2}{r_2 - r_1} \sigma_1 \right)^2 + \left(\frac{r_1}{r_2 - r_1} \sigma_2 \right)^2 + \sigma_{ms}^2 \tag{4.1}$$

where r_1, r_2 denote the radii of the detector layers from the beamline, while σ_1, σ_2 their resolutions. The σ_{ms} term is included to describe the degradation of the resolution due to the multiple scattering of particles as they traverse the detector. It is immediately manifest that the radius of the first detector layer r_1 should be minimised in order to improve the resolution. Thus it is customary to place the first detector layer as close to the beampipe as possible, as evidenced by the IDEA detector concept's reduction in the innermost layer's position from 1.7 cm to 1.37 cm. Subsequent detector layers provide additional measurements which are combined into a single *track*.

The process begins with the formation of hits on each detector layer, consisting primarily of identifying pixels where signals are large enough (although this often occurs internally at hardware level) and combining measurements of adjacent pixels in a process described in Chapter 9. Track seeds, which are initial track candidates, are subsequently formed from two or more hits, typically close to the beampipe. Trajectories are then built by iteratively adding hits to track seeds and using techniques like the Kalman filter [54] to update track parameters at each step to minimise goodness of fit metrics like the χ^2, while accounting for uncertainties and track curvature. The fitting procedure is considerably more involved than the straightline fits described in Chapter 10, as the tracks must follow helical trajectories consistent with the magnetic field strength and particle momentum hypotheses. After all hits have been included in the track candidate, the track is refit in order to refine track parameters, typically using the same fitting method as during the track finding.

---

[1] In the absence of a magnetic field, two points suffice to define a straight line trajectory.

While the tracking concepts discussed here are presented through the lens of silicon tracking detectors, the overall trajectory reconstruction procedure applies more broadly. A similar approach can be adapted for track reconstruction in a gaseous detector like IDEA's drift chamber, albeit with some notable differences in hit reconstruction and resolution.

4.1.1 Vertexing

Tracks are reconstructed into points from which they originate, referred to as *vertices*. The most important of these is arguably the primary vertex, corresponding to the location of the hard scattering, and thus the origin of most tracks associated with its decay products. Particles with lifetimes of $\mathcal{O}(10^{-12})$ s can travel several mm before decaying, leaving behind two or more tracks that are displaced from the primary vertex. These displaced vertices, dubbed *secondary vertices*, are typical of bottom or charmed hadrons, as well as tau leptons, and thus instrumental in their identification. Figure 4.1 shows a schematic of a 3-jet event, where the heavy flavour jet contains a secondary vertex, characterising it distinctly with respect to the other two. The inner component of the tracking detector is often optimized for vertex reconstruction and is thus dubbed the *vertexing detector*.

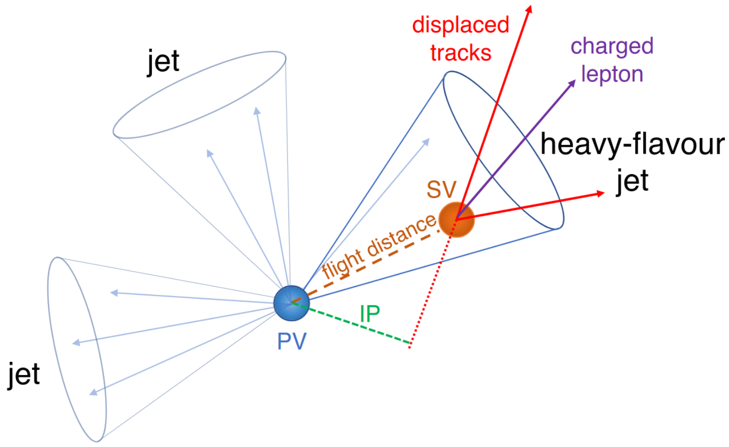


Figure 4.1: Schematic of a heavy flavour jet containing a secondary vertex. The impact parameter of one of the displaced tracks in marked in green. Figure taken from Ref. [55].

The LCFIPlus framework [56], used for vertex reconstruction in the studies presented in Part II of this thesis, employs a three step approach. In the first step, the primary vertex is reconstructed from all tracks in an event, for which a global χ^2 value is computed. Tracks contributing the most to the χ^2 are then iteratively removed until a threshold is reached. The remaining tracks are then assigned to the primary vertex. Subsequently, all possible track pairs constructed from tracks not associated with the primary vertex are considered secondary vertex seeds. Tracks are then added to track pairs with a low χ^2, provided their contribution to the resulting χ^2 remains below a predefined threshold. The procedure continues until all tracks satisfying the criteria given in Ref. [56] have been considered. In the final step of the vertex fitting procedure, V^0 candidates are identified and removed. The studies presented in Chapter 6 deliberately added V^0 candidates as reconstructed objects. In the studies presented in Chapter 6, V^0 candidates were deliberately retained as separate reconstructed objects.

While secondary vertexing performance depends on the precise association of tracks to displaced decay points, the resolution of a track's distance from the primary vertex is often used as a proxy. This is commonly parametrised by the transverse (d_0) and longitudinal (z_0) impact parameters, corresponding to the distance of closest approach in each respective dimension. The performance of a vertexing detector can be quantified via the transverse impact parameter resolution as

$$\sigma_{d_0} = a \oplus \frac{b}{p \cdot \sin^{3/2}\theta} \tag{4.2}$$

where the a and b are parameters describing the *spatial resolution* and *multiple scattering* of a given detector, respectively, while p denotes the particle momentum and θ its angle to the beam axis. The spatial resolution of the sensor is largely driven by its granularity, or so called *pixel pitch* in the case of silicon pixel detectors. Positional reconstruction schemes making reference to the charge deposited in each pixel can be exploited to further improve the spatial resolution of a detector layer, as explored in detail in Chapter 10 for the chips presented in this thesis.

4.2 Particle Reconstruction

The particle flow algorithm [57] has emerged as a powerful paradigm in event reconstruction, combining information from all subdetectors into a single coherent picture. Inputs to the particle flow algorithm include tracks reconstructed from charged particle depositions in the tracking system, clusters in the electromagnetic and hadronic calorimeters formed from energy deposits in neighbouring calorimeter cells, and standalone tracks in the muon system reconstructed from hits in the muon chambers. By combining different signatures into physics objects with corresponding identity and 4-momenta, dubbed particle flow candidates, the algorithm achieves a superior energy resolution for composite objects like jets, and an improved reconstruction accuracy compared to what any individual subdetector can achieve. A particle flow approach is foreseen for the reconstruction of physics objects at the FCC-ee. The Pandora particle flow algorithm [58], for instance, is already well-integrated with the CLD and ILD detector concepts, stemming from its long use in linear collider contexts. In the following, the reconstruction of the particle flow candidates will be briefly described, without reference to a specific detector concept.

4.2.1 Leptons and Photons

Muons are often the first physics object to be reconstructed in particle flow reconstruction chains due to their clean experimental signature. Given that muons do not interact strongly, and are much more massive than electrons, they traverse considerably more detector material before suffering significant energy loss. This is exploited by situating the muon system outside of the superconducting coil, often interweaved with the return yoke, as described for several detector concepts in Section 3.5, such that virtually only neutrinos and muons themselves reach the muon system. Particle flow algorithms then begin by reconstructing the muon trajectory from individual deposits in the muon chambers. The track can be extrapolated to the interaction point, and then matched to the tracks constructed as detailed in Section 4.1, often coupled with additional constraints like minimal calorimeter deposits along the extrapolated track. Deposits in the tracking system are often instrumental in increasing the track resolution of low-energy muons, as these are more subject to multiple scattering. Once a muon particle flow candidate is added to the collection, all associated deposits (e.g. the track) are removed from consideration for subsequent stages of the particle flow algorithm.

The reconstruction of *electrons*, by comparison, is more challenging, and makes use of both tracking information and deposits in the electromagnetic calorimeter. Calorimeter clusters with identifiable tracks along the extrapolated electron trajectory are identified as electrons. Due to substantial bremsstrahlung and multiple scattering, electron tracks are often constructed in a dedicated step with considerably more relaxed criteria than that of muons or charged hadrons. Additional criteria are imposed, including the absence of deposits in the hadronic calorimeter to minimise misidentification from hadrons, and isolation criteria to reject electrons from photon conversions in the detector material. The identification of *photons* is of course intimately related. Bremsstrahlung photons originating from identified electrons are typically not reconstructed as independent objects, but are instead associated with the electron through extended calorimeter clustering. This association ensures that ECAL deposits are correctly attributed and prevents double counting. Clusters that remain unassociated with any track are identified as photons, with the photon energy defined by the ECAL deposit itself.

4.2.2 Hadrons

Unlike the leptons and photons reconstructed above, the identification of hadrons hinges primarily on the absence of a signature. Charged tracks that are not associated to a lepton are considered to originate from charged hadrons, primarily pions ($\pi^{\pm}$), protons (p), and Kaons ($K^{\pm}$). Associated deposits in the hadron calorimeter are used for consistency checks, but are generally considered too coarse to resolve the track invariant mass. Dedicated particle identification techniques exist in the form of ionisation energy loss measurements (dE/dx) or time-of-flight measurements. In either case, the dependence of an uncorrelated observable is coupled with momentum measurements of the tracking system in order to determine the particle mass. These two types of particle identification measurements are particularly effective when combined, as foreseen in the IDEA detector concept, as they cover slightly different momentum ranges, with time-of-flight measurements providing better sensitivity at low momenta [59].

After the subtraction of deposits in the hadronic calorimeter associated with charged hadrons, the remaining deposits are interpreted as originating from *neutral hadrons*. The energy of neutral hadron candidates is defined by the calorimeter deposit, resulting in a considerably worse energy resolution than for other reconstructed particle flow objects. The momentum vector is then defined by the relative position of the calorimeter cluster and a mass hypothesis. Despite the limited precision of neutral hadron reconstruction, it plays an important role in the completeness of the event reconstruction, particularly in observables such as the missing transverse energy.

4.3 Jets

At high energy scales partons produced in a hard scattering emit radiation in the form of gluons as they travel outward, which may themselves branch into quark-antiquark pairs, or further gluons. The cascade of emissions, known as a parton shower, is governed by perturbative QCD and continues until the energy scale approaches $\mathcal{O}(1)$ GeV, where the strong coupling α_s becomes large, and the perturbative treatment of partons begins to break down. In a non-perturbative process known as *hadronisation*, the outgoing partons recombine into colourless hadrons. The exact dynamics of hadronisation are poorly understood, and it is customary to instead resort to phenomenological models, such as the Lund String model [60]. The curious reader is referred elsewhere for the details, but it suffices to note that the model revolves around the treatment of a linear potential between the partons $V(r) \sim k \cdot r$, where k is known as the string tension. In this picture, colour field lines are confined to a narrow tube, analogous to a 1-d electric potential, rather than spreading spherically. As the quarks begin to separate and their potential increases, it becomes energetically favourable for the string to "snap" and produce a quark-antiquark pair from the vacuum. This process repeats iteratively until all partons recombine into colourless hadrons, and the resulting experimental signature is a collimated spray of particles.

Formally a jet is defined with respect to a *jet algorithm*, in other words a mapping from the (hadronic) final state **P** to the *jet pattern* **Q** [61]

$$\mathbf{P} = \{p_1^\mu, ..., p_n^\mu\} \rightarrow \mathbf{Q} = \{q_1^\mu, ..., q_m^\mu\} \tag{4.3}$$

where the p_i^μ and q_j^μ denote the momenta of the i^{th} particle and j^{th} jet in the final state. A *particle* in this thesis refers to a particle-flow object as defined in Section 4.2, unless stated otherwise. Jet clustering algorithms are generally sub-divided into *cone algorithms* and *sequential recombination algorithms*. Cone algorithms exploit the fact that jets appear as flows of energy in conical shapes to define angular regions in Δy and $\Delta\phi$ within which particles are clustered into jets. However, cone algorithms are generally not infrared and collinear (IRC) safe, and have thus largely been superseded by sequential algorithms.

Sequential clustering algorithms, which can be designed to be IRC safe, use less obvious measures often related to the momentum of particle pairs to group particles into pseudo-particles until some minimal distance between pseudo-particles is obtained. The resulting set of pseudo-particles define the jets. The anti-k_T algorithm [62] is defined as

1. Define the clustering set C as the set of all particles in an event
2. Begin by defining the distances between all pairs of (pseudo-)particles $\{d_{ij}|\, i,j \in C \text{ and } i \neq j\}$, and (pseudo-)particles and the beam $\{d_{iB}|\, i \in C\}$
   - $d_{ij} = min(k_{ti}^{-2}, k_{tj}^{-2})\frac{\Delta_{ij}^2}{R^2}$
   - $d_{iB} = k_{ti}^{-2}$
3. Find $min(\{d_{ij}|\, i,j \in C \text{ and } i \neq j\},\, \{d_{iB}|\, i \in C\})$
   - if $min(...) \in \{d_{ij}|\, i,j \in C \text{ and } i \neq j\}$, combine (pseudo-)particles i and j into new pseudo-particle, and remove (pseudo-)particles i and j from C
   - if $min(...) \in \{d_{iB}|\, i \in C\}$, define (pseudo-)particle i as a jet and remove (pseudo-)particle i from the C
4. Repeat from 2. until there are no (pseudo)particles left

where the distance measure $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is defined as the $(y - \phi)$ separation between (pseudo)particles i and j. The anti-k_T algorithm is known to result is roughly conical-shaped jets, as depicted on Figure 4.2 where it is shown alongside a number of other clustering algorithms, whose size is defined by the parameter R.

Although the anti-k_T algorithm is widely adopted at the LHC, it is not very well suited for e^+e^- collisions where spherical coordinates provide better measures of angular distance. The

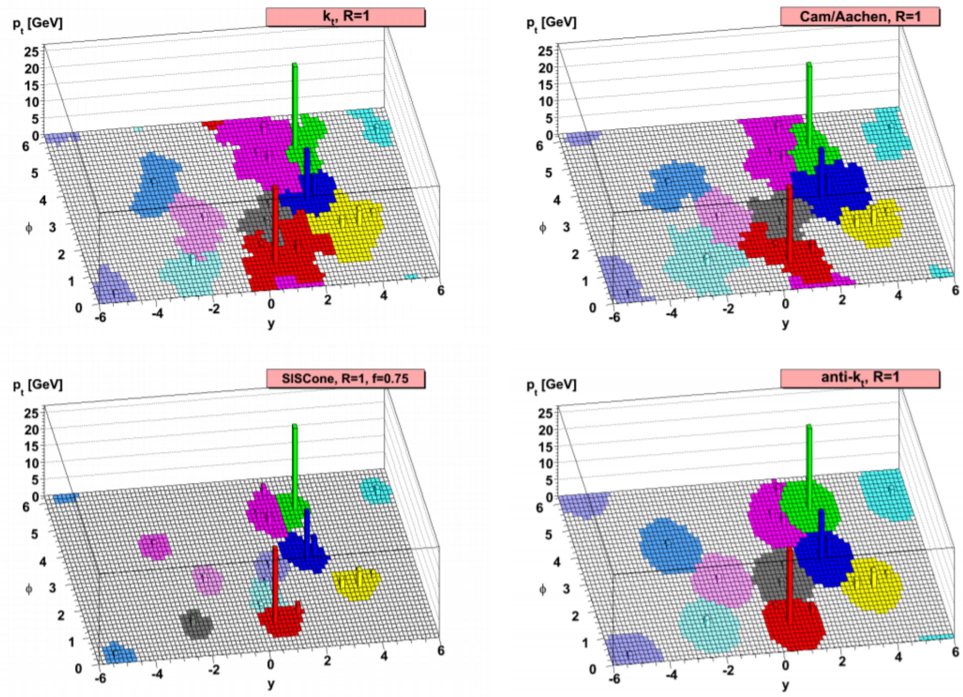


Figure 4.2: Jets clustered from partons with the k_t, Cambridge/Aachen, SISCone, and anti-k_t algorithms. Soft ghosts are used to indicate catchment areas. Figure taken from Ref. [62].

generalised e$^+$e$^-$ k_T algorithm [63] replaces the distances d_{ij} and d_{iB} by

- $d_{ij} = min(E_i^{2p}, E_j^{2p})\frac{(1-\cos(\theta_{ij}))}{(1-\cos(R))}$
- $d_{iB} = E_i^{2p}$

where θ_{ij} is the angle between particles i and j, and p, R are parameters of the algorithm. In processes where a certain number of jets is expected, such as the $Z \to q\bar{q}$ decays studied in Chapter 6, the clustering recipe can be modified to stop once a certain number of jets is reached. This is the case for the e$^+$e$^-$ k_T, or Durham, algorithm [64], which only includes a single distance metric d_{ij}

- $d_{ij} = 2\, min(E_i^{2p}, E_j^{2p})\,(1-\cos(\theta_{ij}))$

and was adopted for the results presented in Part II of this thesis.

While the jet clustering algorithm defines the mapping presented in Equation 4.3, jets are often treated as proxies to the final state partons originating from the hard scattering. Considering the notion of jets as collimated sprays of particles resulting from the hadronisation of a given parton, the *jet flavour* can loosely be defined as the flavour of the initiating parton. The notion is ill-defined since a jet is not guaranteed to exclusively include the particles resulting from the hadronisation of a single parton, or indeed only particles from the hard scattering. This may at first seem salvageable by noting that the ambiguity arises from imperfect knowledge of the event, and that if each hadron were perfectly reconstructed and mapped back to its initiating parton, then the flavour definition given above would hold. However, the definition still depends on the notion of a *final state parton*, which is inherently energy scale dependent, as gluon emissions alter the final state structure. Even the notion of a parton is only well-defined at tree level, beyond which the concept becomes increasingly ambiguous. Thus it becomes necessary to approach the problem of jet flavour in the same way as the problem of jet clustering was approached above: with a precise recipe.

The *physics definition*[2] derives its name from the intuitive picture above, and matches the flavour of the jet to the flavour of the parton from which it was initiated. The ambiguities discussed above are dealt with by constraining this definition to the leading order contribution to a scattering process. The physics definition was adopted for the studies presented in Part II of this thesis, where the $Z \to q\bar{q}$ and $Z(\to \nu\nu)H(\to q\bar{q})$ events allow the assignment of jet flavour without reference to a particular jet. Generalising the definition to more non-trivial decay topologies is done by matching the parton lying within $R < 0.3$ of the given jet's axis in $(\eta - \phi)$ space. Jets for which this not is unambiguously possible are left undefined. While generally superseded by the ghost matching definition presented below at CMS, the physics definition is well-motivated at lepton colliders, where the parton to jet matching is considerably more straightforward.

The *ghost-matching* definition [66] revolves around the inclusion of hadrons containing b or c quarks, herein referred to as b and c hadrons, during a separate reclustering of the jets. In particular the process begins with the rescaling of the momenta for b and c hadrons that do not have b and c hadrons as daughter particles, as well as partons from the end of the parton showering, by a small constant $\epsilon = 10^{-18}$, turning them into "ghost". Jets are then reclustered with the inclusion of ghosts, which do not affect the jet kinematics. Flavour is then assigned according to the presence of a ghost b or c hadron, with precedence given to b hadrons, such that jets containing both are considered b-jets. Jet flavour for jets containing neither a b nor c hadron is further subdivided with reference to the clustered partons. If the jet contains a b or c parton then it is assigned the respective flavour in analogy to above. For jets containing neither, the hardest light quark or gluon defines the jet flavour.

Having surveyed how a jet is constructed and its flavour defined, the following part of this thesis explores how to map the spray of particles back to its assigned jet flavour. Chapter 5 begins by introducing a powerful tool that has emerged over the past few decades: Machine Learning. Given the extensive availability of labelled datasets in the form of Monte Carlo simulation, machine learning has emerged as the *de facto* avenue to exploiting the subtle correlations characteristic of hadronisation.

---

[2] The jet flavour definitions presented here stem from conventions of the CMS experiment [65].

Part II

Jet Flavour Tagging at the FCC-ee

5 Machine Learning

Machine learning (ML) broadly describes the emergence of algorithms from specified boundary conditions and learning rules, rather than being explicitly programmed. The concept revolves around the iterative improvement of algorithms as they acquire more experience, often with minimal human intervention, standing in analogy to Darwinian evolution. Ref. [67] provides a definition for ML as

> "A computer program is said to *learn* from experience **E** with respect to a class of tasks **T** and performance **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**."

where the task **T** in Ref. [67] is playing checkers, while the performance **P** is the probability of winning a game, and the experience **E** is the amount of games played.

Machine learning techniques are largely subdivided along the **T**/**E**/**P** into three categories: supervised learning, unsupervised learning, and reinforcement learning. In *supervised learning* the task revolves around learning a mapping from a set of inputs $\{\vec{x}\}$ to a set of outputs $\{\vec{y}\}$. In the case where the outputs $\{\vec{y}\}$ are classes, the task is referred to as classification, while in the case where the outputs are real-valued, the task is referred to as regression. The experience takes the form of labelled inputs, turning the problem into a minimisation problem according to some metric defined on the labels of the given inputs, and the prediction of the algorithm. These are referred to as *loss-functions*, and are detailed in Section 5.2. In *unsupervised learning* no such labels exist, and the task is more abstractly to learn latent structure in the input data. Classic examples of unsupervised learning include clustering, dimensionality reduction, and anomaly detection. One of the primary challenges in unsupervised learning lies in defining a meaningful measure of performance, which is often task-dependent. For clustering, for instance, this could be the silhouette score, which revolves around the distance of each data point to its assigned cluster compared to other clusters. In *reinforcement learning* the task revolves around decision-making in some environment. Examples include decisions in a game, robotics, and self-driving cars. While often considered one of the most promising areas of machine learning, reinforcement learning also presents unique challenges due to the need for exploration and delayed reward signals. The experience is not neatly available in the form of data, but rather is acquired through interaction with its environment. The performance is then defined according to a reward function that reflects the desired outcome following the actions taken. The goal is to maximise the cumulative reward through interaction with the environment.

All three types of machine learning have been gainfully applied in high-energy physics, with early applications dating as far back as the 1980s. Supervised learning has been applied to classification tasks including particle identification [68] and jet tagging [69]. More generally, supervised learning has been employed in the classification of events, including BSM particles [70]. Regression-based calibration tasks involve the use of machine learning models to predict the values of true physical observables based on reconstructed detector data. These include jet energy calibration [71] and detector response calibration [72]. Two tasks that have greatly benefitted from unsupervised learning are anomaly detection [73] and generative modelling [74], where explicit labelling is either not possible or impractical. The identification of signals deviating from the Standard Model background relies on the model's ability to detect statistical anomalies without prior knowledge of the signal's characteristics. In generative modelling, the underlying distribution of the physics data is learned, allowing for fast sampling of distributions that are numerically challenging. This manifests as learned simulations that are substantially faster than their traditional MC counterparts. Reinforcement learning is arguably the least explored of the three within the field of HEP, not least due to the computational challenges it entails. A promising application is beam control at collider experiments, where precise tuning of parameters is required in order to ensure optimal beam parameters, including luminosity and stability. Recent attempts [75] at autonomous control of a particle accelerator have demonstrated a better-than-human performance in beam quality in high-fidelity simulation.

Given that the work in this thesis revolves around the application of a transformer architecture to the classification of jets at the FCC-ee, focus will be given to modern deep learning techniques applied to supervised learning. In particular, this chapter provides the necessary foundations to understand the results presented in Chapter 6, and the architecture of *DeepJetTransformer* [2] detailed in Section 6.3. The chapter begins by delving into the foundational block of deep learning: the fully-connected neural network. Techniques for training of neural networks are presented, along with challenges often encountered. Subsequently, more sophisticated architectures are presented, with special emphasis on those appearing in DeepJetTransformer. Finally, the core building block of DeepJetTransformer is presented: the Self-Attention Mechanism. The chapter concludes with a brief overview of machine learning in jet flavour tagging at the LHC.

5.1 Neural Networks

A natural starting point for any discussion on machine learning is the introduction of the perceptron [76] by Frank Rosenblatt. The idea was to mimick the functioning of a neuron by connecting information encoded in a set of features $\{\vec{x}\}$, in order to produce a single output $y = \pm 1$, corresponding to whether or not the neuron had fired. The remarkably simple algorithm consisted of a linear combination of the input features

$$y = sign(w_0 + w_1 x_1 + ... + w_n x_n) \tag{5.1}$$

where the w_i are the trainable weights of the algorithm. The update rule of the perceptron revolves around adding the input vector of any incorrect prediction y_t to the weight vector as

$$\vec{w} \rightarrow \vec{w} + y'_t \cdot \vec{x}_t \tag{5.2}$$

where y'_t corresponds to the correct prediction for input vector x_t. It was proven that the perceptron converges for a finite number of passes [77]. Nevertheless, the perceptron is constrained to linearly separable problems. Extending the perceptron to non-linear problems requires introducing non-linearities into the algorithm. The first approach revolves around extending the basis of the input vectors with non-linear feature transformations. This involves introducing a mapping to a higher-dimensional space, where the data does become linearly separable. A conceptually simple yet powerful approach is extending the basis to include polynomial terms, which for a simple 2-d vector becomes

$$(x_1, x_2) \to (x_1, x_2, x_1^2, x_2^2, x_1x_2, ...) \tag{5.3}$$

It quickly becomes clear that while a polynomial basis expansion is much more expressive than a linear model, it suffers from an exponential increase in basis dimension if the data lies in a high-dimensional space. This is commonly referred to as the *curse of dimensionality*. Kernel methods provide a way to address this challenge by implicitly mapping data to high-dimensional spaces using kernel functions, avoiding explicit feature computation. Nevertheless, kernel methods still suffer from computational complexity, as large datasets require computing the $N \times N$ kernel matrix, which can become prohibitively expensive.

An alternative approach to increasing the expressiveness of the perceptron revolves around noting that multiple perceptron outputs can be concatenated into an output vector $\vec{y}$, where each of the elements of the output vector is referred to as a neuron:

$$(x_0, ..., x_n) \to (\sum_i w_{0i}x_i, ..., \sum_i w_{mi}x_i) = (y_0, ..., y_m) \tag{5.4}$$

where the w_{ij} are the weights connecting input j to output i. Clearly, the operation could be repeated on the output vector $\vec{y}$, but this would not make the model more expressive, since a linear combination of linear combinations is itself a linear combination of the input vector. Non-linearity is introduced to the model by subjecting the output neurons of each linear layer given in Equation 5.4 to a non-linear function f, denoted the activation function. Including this intermediate step in the computation of $\vec{y}$ gives

$$(x_0, ..., x_n) \to (f(\sum_i w_{0i}x_i + b_0), ..., f(\sum_i w_{mi}x_i + b_m)) = (y_0, ..., y_m) \tag{5.5}$$

where bias terms b_i are included to shift the activation function. The above operation, consisting of a linear combination followed by a non-linear activation function applied element-wise, is referred to as a fully-connected layer.

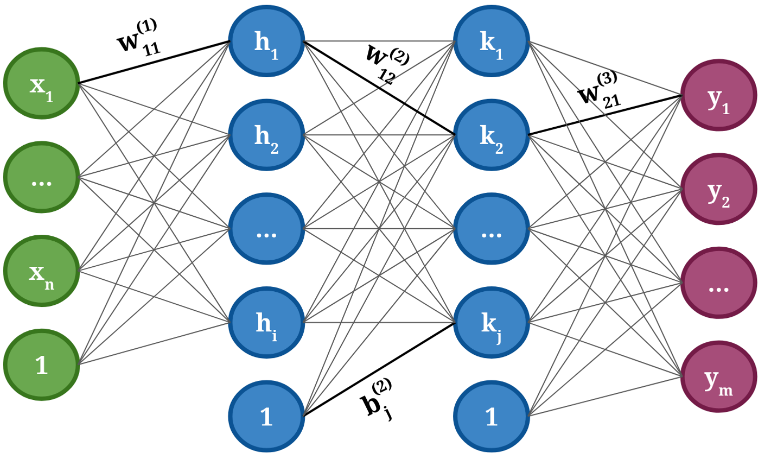


Figure 5.1: Simplified schematic of a 4-layer MLP. Edges denote weighted addition. Activation functions are not explicitly depicted. Throughout this section $n = 100$, $i = j = 512$, $m = 10$ are assumed.

One of the most common examples of an activation function is the rectified linear unit (ReLU), defined as

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{5.6}$$

Another widely used activation function is the *softmax* function, which is defined over a vector of inputs $\vec{v} = \{v_0, ..., v_n\}^\intercal$ as

$$f(\vec{v})_i = \frac{\mathrm{e}^{v_i}}{\sum_l \mathrm{e}^{v_l}} \tag{5.7}$$

and is often used to map the output vector of a neural network to a vector of numbers in the interval $(0, 1)$, which sum to 1, allowing the interpretation of the vector as a probability distribution. The ReLU activation function is typically favoured for hidden layers, due to its relative simplicity and piecewise constant derivative, while the softmax function often appears at the very end of the network, particularly during inference.

The seemingly trivial adjustment of introducing an activation function has profound implications, by stacking such layers a Multi-Layer Perceptron (MLP) is obtained:

$$\vec{x} \to h_1(\vec{x}) \to h_2 h_1(\vec{x}) \to h_n ... h_1(\vec{x}) \to \vec{y} \tag{5.8}$$

where $\vec{x}$, $\vec{y}$ are the input and output vectors respectively, and h_i is a fully-connected layer given by Equation 5.5. The Multi-Layer Perceptron was proven to approximate any "well-behaved" function arbitrarily [78]. The MLP laid the foundations of Deep Neural Networks (DNNs), with early theoretical results showing their universal approximation capability, albeit limited by training and hardware bottlenecks.

Nevertheless, DNNs are notoriously compute-heavy. It suffices to consider the number of multiplications in a forward pass for a 4-layer network presented in Figure 5.1:

$$N_{mult.} = \sum_{i=1}^{n_{layers}} N_i \cdot N_{i-1} = 100 \cdot 512 + 512^2 + 512 \cdot 10 = 318464 \tag{5.9}$$

where N_i is the number of neurons in the i^{th} layer with N_0 being equal to the input dimension. Even a comparatively simple MLP presents a formidable challenge for the Von Neumann CPU architecture, which requires inputs to be processed sequentially at a high clock frequency, making such a large number of multiplications, and their associated memory lookups, quickly intractable. It was not until the advent of heterogeneous computing, and the Graphics Processing Unit (GPU), that interest in DNNs resurged.

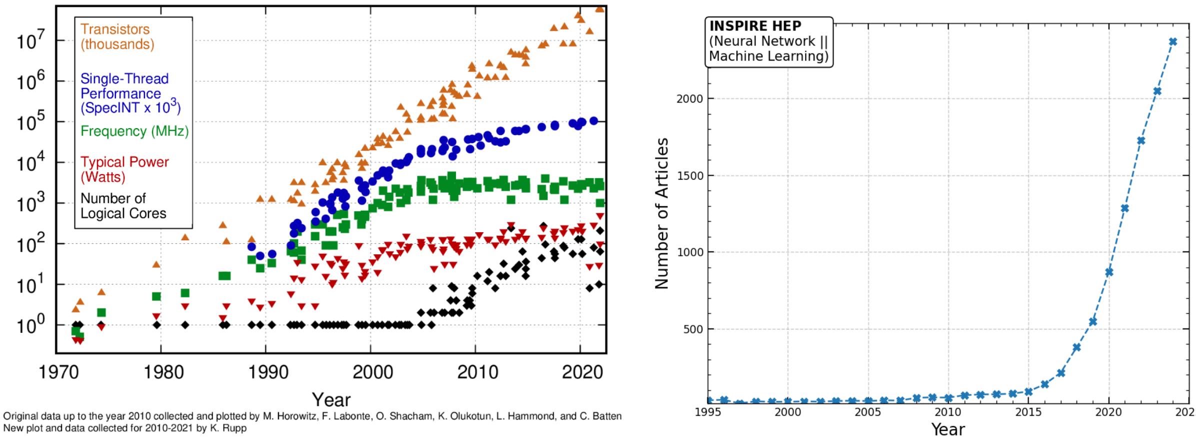
GPU Paradigm

Indeed, Equation 5.5 can be rewritten as a matrix multiplication by defining the input vector as $\vec{x} = (x_0, ..., x_n)^T$

$$\vec{x} \to \mathbf{W} \cdot \vec{x} + \vec{b} \to f(\mathbf{W} \cdot \vec{x} + \vec{b}) \to \vec{y} \tag{5.10}$$

where the $\mathbf{W}$ is the weight matrix whose elements w_{ij} are given as in Equation 5.5, $\vec{b} = (b_0, ..., b_m)^T$ is the bias vector, $\vec{y} = (y_0, ..., y_m)^T$ is the output vector, and f is the activation function. Not only is the matrix form of Equation 5.5 considerably more concise, but it highlights that the bulk of the computation in a neural network consists of matrix multiplications. As ubiquitous as matrix multiplications are, one field of computing that saw a particular need for their optimisation was computer graphics. In computer graphics, operations such as rotations, scaling, and projecting 3d objects onto 2d space, are implemented as repeated matrix multiplications. The acceleration of these tasks hinged on the development of GPUs, which unlike their CPU counterparts feature a lower per-thread performance, but typically include thousands of threads. The slowdown in the rate of improvement of the single-thread performance of CPUs in the early 2000s, depicted on Figure 5.2a, motivated the adoption of parallel architectures. The high number of threads that GPUs feature is uniquely suited for parallelisable tasks, including matrix multiplication, where concurrent execution can effectively hide the latency of the computation.

Thus the wider adoption of DNNs can be regarded as having been hardware-gated [79], with the emergence of programmable consumer-grade GPUs in the 2000s effectively enabling their training, for the first time. While shallow neural networks could be trained on CPUs, the introduction of GPU-based training in the mid-2000's [80, 81], and the demonstration of its performance on large-scale tasks [82], marked the beginning of modern deep learning.



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp

Figure 5.2: Overview of transistor count, single-thread performance[1], clock frequency, power use, and core count of flagship CPUs from major vendors since the 70s. Figure adapted from Ref. [83] (a). Number of articles with keywords "Neural Network" or "Machine Learning" by year in INSPIRE-HEP since 1995 (b).

Figure 5.2b depicts the number of INSPIRE-HEP articles containing the keywords "Neural Network" or "Machine Learning". While GPU-training of neural networks had already begun in the mid-2000's, it was not until the advent of deep learning libraries such as Tensorflow [84] and PyTorch [85] around 2015, that deep learning became widespread in HEP. These libraries abstracted the often cumbersome and error-prone process that is writing GPU kernel functions, even compared to earlier frameworks that were primitive and lacked robust GPU support.

5.2 Learning as Optimisation

The goal of training a neural network is the minimisation of the loss-function particular to the given task with respect to the network parameters. In the case of classification, a commonly used loss function is the categorical cross-entropy given by

---

[1] Based on SPEC CPU benchmark data. See `https://www.spec.org/cpu/`.

$$L(\vec{t}, \vec{y}) = -\sum_{i}^{N_c} t_i log(y_i) \tag{5.11}$$

where $\vec{t}$ refers to the truth vector, defining the class to which a given input belongs, and $\vec{y}$ is the vector of predictions given by the neural network.

Denoting the weights of a single output neuron for a single-layer neural network as $\vec{w}$ gives

$$\vec{w}^* = \arg\min_{\vec{w}} L(\vec{w}) \tag{5.12}$$

where $\vec{w}^*$ are the optimal weights. For a given weight $w_i = (\vec{w})_i$ we wish to find the roots of the first derivative, which we can do by performing a Taylor expansion about some initial estimate w_i^0:

$$\begin{aligned}
\frac{\partial L(w_i)}{\partial w_i} &= \frac{\partial}{\partial w_i}\left(L(w_i^0) + \left.\frac{\partial L(w_i^0)}{\partial w_i}\right|_{w_i=w_i^0}(w_i - w_i^0) + \frac{1}{2}\left.\frac{\partial^2 L(w_i^0)}{\partial w_i^2}\right|_{w_i=w_i^0}(w_i - w_i^0)^2 + \cdots\right) \\
&\approx \frac{\partial}{\partial w_i}\left(L(w_i^0) + \left.\frac{\partial L(w_i^0)}{\partial w_i}\right|_{w_i=w_i^0}(w_i - w_i^0) + \frac{1}{2}\left.\frac{\partial^2 L(w_i^0)}{\partial w_i^2}\right|_{w_i=w_i^0}(w_i - w_i^0)^2\right) \\
&= \left.\frac{\partial L(w_i^0)}{\partial w_i}\right|_{w_i=w_i^0} + \left.\frac{\partial^2 L(w_i^0)}{\partial w_i^2}\right|_{w_i=w_i^0}(w_i - w_i^0)
\end{aligned} \tag{5.13}$$

Setting the LHS equal to 0, gives

$$w_i = w_i^0 - \frac{L'(w_i^0)}{L''(w_i^0)} \tag{5.14}$$

which is the well-known Newton's method for optimisation, typically expressed as an update-rule with $w_i^{n+1} = w_i^n - \frac{L'(w_i^n)}{L''(w_i^n)}$. As a second order method, the algorithm converges quadratically, with the L' term giving the direction and magnitude of the function's growth, while the L'' term accounts for the local curvature. Generalising the method to higher dimensions, we obtain

$$\vec{w}_{n+1} = \vec{w}_n - \left[\nabla^2 L(\vec{w}_n)\right]^{-1} \nabla L(\vec{w}_n) \tag{5.15}$$

where $\nabla^2 L(\vec{w}_n)$ is the Hessian matrix at $\vec{w}_n$, and the subscripts no longer denote the component of the vector, but the step in the iteration. However, Equation 5.15 reveals two aspects of Newton's method that have prevented widespread adoption when training neural networks: computational complexity and numerical instability. In particular, the Hessian is a $N \times N$ matrix of the mixed partial derivatives that must be inverted, which can be computationally expensive for large numbers of weights. Moreover, if the Hessian is singular or ill-conditioned then the update of $\vec{w}_{n+1}$ can either not be performed, or leads to numerical instability by distorting the update vector.

This can be addressed by simply replacing the Hessian with a constant η^{-1}, such that the update rule becomes:

$$\vec{w}_{n+1} = \vec{w}_n - \eta \cdot \nabla L(\vec{w}_n) \tag{5.16}$$

and is denoted as the *gradient descent.* Although gradient descent converges only linearly as it is a first-order method, it circumvents the computational challenges that the Hessian introduces. Gradient descent became the de facto foundation for neural network training algorithms following its successful application in 1986 by Rumelhart, Hinton, and Williams [86].

Computing the gradient of the loss over all examples in the dataset before updating the weights, known as *batch* gradient descent, is computationally expensive for large datasets. In *stochastic* gradient descent the gradient of the loss is computed for individual training examples before updating the weights according to Equation 5.16. While the advantage is immediately obvious: reduced computation, this also entails a noisier training which may not converge smoothly. Thus *mini-batch* gradient descent offers a compromise, where the derivative of the loss is computed over a small sample of size n before updating the weights. Typical values for the batch size n are $32, 64, 128, ...$, balancing a quick computation of the gradients with reduced noise resulting from the averaging.

Modern Optimisers

Despite offering a simple recipe for weight optimisation, gradient descent often struggles in regions of high curvature, leading to slow convergence or oscillatory behaviour. Momentum methods were introduced to account for the value of previous gradients:

$$v_n = \beta \cdot v_{n-1} + (1 - \beta) \cdot \nabla L(\vec{w}_n) \tag{5.17}$$

where v_t is denoted as the velocity, which can be interpreted as a smoothed gradient, and $\beta \in [0, 1)$ is a parameter controlling the extent of the smoothing. The update rule then takes the form

$$\vec{w}_{n+1} = \vec{w}_n - \eta \cdot v_n \tag{5.18}$$

Extensions include the Adam optimiser [87], which in addition to the velocity term that exponentially averages the gradient, includes an uncentered variance term that exponentially averages the square of the gradient. Adopting a convention where the velocity is denoted by m_n and the variance by v_n, the two terms are given by

$$m_n = \beta_1 \cdot m_{n-1} + (1 - \beta_1) \cdot \nabla L(\vec{w}_n) \tag{5.19}$$

$$v_n = \beta_2 \cdot v_{n-1} + (1 - \beta_2) \cdot (\nabla L(\vec{w}_n))^2 \tag{5.20}$$

where the square in the computation of v_n applies component-wise, and the $\beta_{1,2} \in [0, 1)$ parameters control the degree of the exponential averaging for the velocity m_n and variance v_n, respectively. Since the velocity and variance are initialised at 0, this can lead to excessively small updates at early training steps. The terms are bias-corrected by scaling them as

$$\hat{m}_n = \frac{m_n}{1 - (\beta_1)^n}, \quad \hat{v}_n = \frac{v_n}{1 - (\beta_2)^n}$$

where it can be seen that the $(\beta_1)^n$ and $(\beta_2)^n$ terms scale up the velocity and variance early into the training, but vanish at high training epochs. The term $\hat{v}_n$ is then used to scale the velocity in Equation 5.18, accounting for directions in which the gradients have consistently been large and taking smaller steps, leading to the Adam optimiser update rule:

$$\vec{w}_{n+1} = \vec{w}_n - \eta \cdot \frac{\hat{m}_n}{\sqrt{\hat{v}_n} + \epsilon} \tag{5.21}$$

where the ϵ is a small term added for numerical stability. The Adam optimiser tends to struggle with over-aggressive weight updates early in the training process, before the bias-corrected velocity and variance terms stabilise. The rectified Adam optimiser (RAdam) [88] addresses this by introducing a term ρ_n that serves to estimate to the reliability of the variance v_n, and reducing the Adam optimiser update rule to a simple momentum update rule early into training. In particular, the update rule is given by Equation 5.18 if $\rho_n < 5$, as the adaptive learning rate component is considered unreliable below this threshold. The ρ_n term is defined as

$$\rho_n = \rho_\infty - \frac{2n \cdot (\beta_2)^n}{1 - (\beta_2)^n}, \quad \rho_\infty = \frac{2}{1 - \beta_2} - 1$$

For $\rho_n > 4$ a rectification term r_n is introduced, which is defined as

$$r_n = \sqrt{\frac{(\rho_n - 4)(\rho_n - 2)\rho_\infty}{(\rho_\infty - 4)(\rho_\infty - 2)\rho_n}} \tag{5.22}$$

The rectification term is included in the Adam optimiser update rule to suppress updates early into the training, and tends towards 1 as $n \to \infty$:

$$\vec{w}_{n+1} = \vec{w}_n - r_n \cdot \eta \cdot \frac{\hat{m}_n}{\sqrt{\hat{v}_n} + \epsilon} \tag{5.23}$$

Having detailed some sophisticated update rules for the weights, attention can be shifted to how to train the full neural network.

Back-propagation

Until now we have constrained our discussion only to the optimisation of the weights in a single layer of a neural network. While instructive, this is clearly not the full story, and it is evident that the weights appearing in the final layer of a neural network cannot be optimised without considering those before it, complicating the problem. Rumelhart, Hinton, and Williams offered the solution in their landmark 1986 paper [86]. Their landmark paper introduced the concept of back-propagation, whereby the weights of the final fully-connected layer are related to weights of earlier layers by the chain rule. Let $\vec{x}^n$ denote the output neurons of layer n, then for the last layer we have

$$\frac{\partial L}{\partial w_{ij}^n} = \sum_l \frac{\partial L}{\partial y_l} \cdot \frac{\partial y_l}{\partial w_{ij}^n} = \frac{\partial L}{\partial y_j} \cdot \frac{\partial y_j}{\partial w_{ij}^n} \tag{5.24}$$

where y_j denotes the j^{th} neuron of the output layer and w_{ij}^n denotes the weight connecting it to x_i^n. Considering the partial derivative of the loss-function with respect to a weight w_{ki}^{n-1} of the previous layer gives

$$\frac{\partial L}{\partial w_{ki}^{n-1}} = \frac{\partial L}{\partial x_i^n} \cdot \frac{\partial x_i^n}{\partial w_{ki}^{n-1}} = \sum_l \frac{\partial L}{\partial y_l} \cdot \frac{\partial y_l}{\partial x_i^n} \cdot \frac{\partial x_i^n}{\partial w_{ki}^{n-1}} \tag{5.25}$$

This procedure generalises naturally to earlier layers, where the partial derivative of each weight w_{ki}^{l-1} can be computed by evaluating $\frac{\partial L}{\partial x_i^l}$ recursively. In this way, the gradients are propagated backward through the network, allowing the computation of the updated weights at each layer according to Equation 5.16.

Training and Performance

The introduction of gradient descent provided a practical algorithm for training the largely theoretical constructs that neural networks had been, while the emergence of parallel hardware made it technically feasible. This led to an explosion in the complexity of models, and with it exacerbated a problem familiar to curve fitting: *overfitting*. The ultimate goal in classification supervised learning, in the words of the performance **P** introduced earlier in this Chapter, is the correct prediction of an arbitrary element of a given class according to its label. While minimising the loss-function according to the training dataset improves the performance of the algorithm as distinguishing features are learned, this does not continue indefinitely, and models often being extracting patterns that are particular to the given dataset, rather than general properties of the labelled classes. Thus one of the most important cornerstones of training a neural network is splitting the dataset into training, validation, and test sets, with a typical split being 80:10:10. The *training* set is the data set on which the loss-function is minimised, comprising the vast majority of the computation as gradients are back-propagated. The *validation* set is kept separate, and no weight-updates are performed with reference to it. Rather, at the end of each training epoch after the training dataset has been traversed, the loss is evaluated on the validation set. This is instrumental for identifying if the model has overfit the training set. Figure 5.3a depicts the loss computed on the training set and the validation sets at the end of each epoch for a typical neural network. The training and validation losses initially decrease very rapidly, before incrementally improving. After some epoch the training and validation losses diverge as the model begins to overfit the training dataset. Thus it is often prudent to include a stopping condition for when the model's performance on the validation set stops improving. Typically stopping conditions include early stopping, whereby training is halted if the validation loss does not improve over a set number of epochs, or a maximum cap is placed on the number of epochs, regardless of convergence. The *test* set is the final set and does not partake in the training of the model. Rather, it is set aside until the hyperparameters and weights of the model have been fixed, and only then is the model evaluated on it. In this way, it is possible to obtain an unbiased estimate for the performance of the model on the task **T**, and its ability to generalise what has been learned on the training dataset.

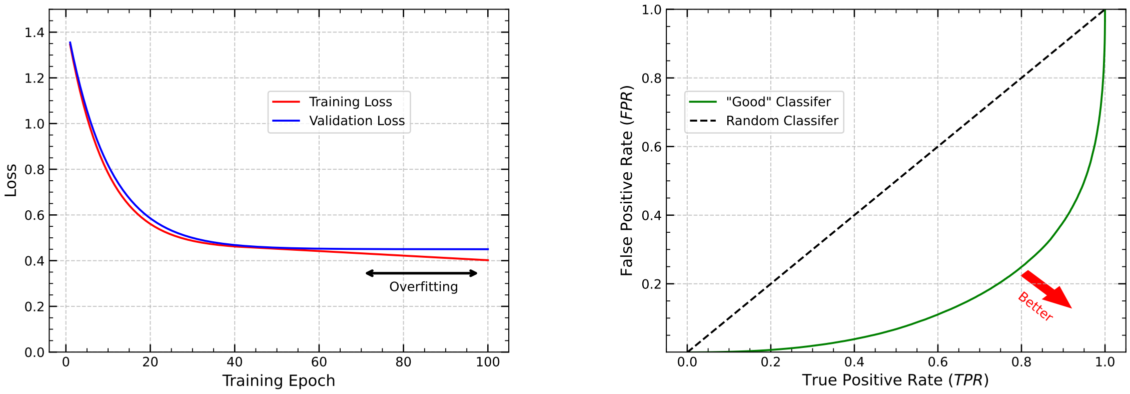


Figure 5.3: Schematic[2] training and validation loss curves as a function of training epoch. At high epochs the training loss continues to decrease, despite the validation loss stagnating, indicating overfitting (a). Typical ROC curve depicting the true positive rate on the x-axis and the false positive rate on the y-axis at varying classifier thresholds. A random classifier is included as the dashed black line for reference (b).

While the loss serves as the arbiter of the model performance during training, it is somewhat abstract, and additional metrics are often included when considering the performance of the model on the test set. In the case of classification, it is useful to describe the rate at which elements of a class are identified correctly and incorrectly. The *true positive rate* describes the fraction of elements of a class c that are correctly identified as belonging to that class, i.e.

$$TPR = \frac{N_{(y^p=c|t=c)}}{N_{(y^p=c|t=c)} + N_{(y^p\neq c|t=c)}} \tag{5.26}$$

where y^p describes the class predicted by the network, and t describes the true class the element belongs to. The network prediction for class c can be obtained by considering the c^{th} component of the output vector $\vec{y}$, appearing for instance in Equation 5.8, and setting a decision threshold τ_c:

$$y^p = \begin{cases} c & \text{if } y_c > \tau_c \\ \neg c & \text{if } y_c \leq \tau_c \end{cases} \tag{5.27}$$

The true positive rate is used as a proxy for the probability of correctly identifying an element of class c: $P(y^p = c|t = c)$. While it is clear that a high true positive rate is desirable, this comes at the cost of incorrectly labelling elements of other classes as belonging to the class c. The "cost" of incorrectly labelling elements of other classes is given by the *false positive rate*

$$FPR = \frac{N_{(y^p=c|t\neq c)}}{N_{(y^p=c|t\neq c)} + N_{(y^p\neq c|t\neq c)}} \tag{5.28}$$

The tradeoff between the true positive rate and the false positive rate is given by computing a *receiver operating characteristic* (ROC) curve, obtained by computing the true- and false positive rates at a number of different decision thresholds τ_c, appearing in Equation 5.27. Figure 5.3b depicts a ROC curve, where the x-axis is the true positive rate, and the y-axis is the false positive rate. Although the more common convention flips the axes, with the true positive rate appearing on the y-axis, Figure 5.3b adopts an axis orientation often used in jet flavour tagging. Classifiers are considered to be more performant the closer the ROC curve is to the bottom-right corner, suggesting a high true positive rate, and a low false negative rate. On the other hand, a random classifier would appear along the $x = y$ diagonal. This corresponds to guessing, with the probability of an element assigned to a class truly belonging to that class being 50% in a balanced binary dataset.

Though the above discussion has been deliberately general with respect to what constitutes the background to class c, it can be subdivided into two evaluation strategies commonly referred to as one-vs-rest and one-vs-all. In the one-vs-rest evaluation scheme the class c is taken to be the signal, while all non-c elements are treated as the background. This entails that the entire dataset is used in the computation of the ROC curve, but also that the ROC curve will depend on the distribution of the classes in the dataset. If for instance a class that is particularly difficult to resolve with respect to c, then the measured performance of the classifier will be worse. On the other hand, in the one-vs-one evaluation scheme, the class c is considered only with respect to one background class b at a time. In other words, the expressions $y^p \neq c$, $t \neq c$ appearing in Equations 5.26 and 5.28 are replaced by $y^p = b$, $t = b$. The one-vs-one evaluation strategy is adopted throughout this thesis, and in particular in Section 6.5 presenting the jet flavour tagging

---

[2] These curves are illustrative and do not correspond to an actual NN training.

results. With the foundations of fully-connected neural networks, along with their training and evaluation, established, the following Section introduces more sophisticated architectures that have shown strong performance in jet flavour discrimination. A key distinguishing feature of these architectures is how jets are represented.

5.3 Beyond Fully Connected Networks

A primary consideration in the architecture of a neural network is the input representation. While the MLPs introduced in Section 5.1 are powerful in their own right, they are not well suited for geometric inputs, such as those arising in jet tagging. Though MLPs can be useful, they typically require extensive pre-processing of the data. High-level features are constructed which capture dependencies between jet constituents explicitly. For instance, in the case of quark gluon discrimination features like the jet mass or the charge-weighted track p_T are constructed to capture the discrepancies expected from the differing colour factors and charges of quarks and gluons [89]. Many applications in machine learning, including HEP, arise from information measured by detectors and sensors. This information often comes coupled with spatial information, and is thus laid out in an irregular cluster of points in space, denoted a *point cloud.* Each point in a point cloud corresponds to a set of information. In the case of jet tagging each point in the cloud might represent a particle, with its location corresponding to kinematic coordinates (e.g. p_x, p_y, p_z), and additional features including the energy, charge, and particle type being treated as attributes of the point. The point cloud representation aligns naturally with the particle-flow reconstruction detailed in Chapter 4, where each candidate is treated as a point by virtue of not only the tracking information, but also the precise location of the calorimeter deposits with respect to the interaction point. While point clouds encode geometric information well, they are not particularly easy to deal with from a neural network perspective due to their irregular, unordered nature. A common approach, and the first that will be discussed in this section, is mapping the point cloud to a regular grid, and subsequently using a *convolutional neural network* (CNN) to extract local features. Figure 5.4a depicts this approach for jet tagging by representing the jet as a set of $|p|$-weighted images binned with respect to distance in $\theta - \phi$ from the jet axis. An alternative, and more direct, approach revolves around treating the point cloud as an unordered graph and using a *graph neural network* (GNN) to process the input, as depicted in Figure 5.4b.

Convolutional Neural Networks

CNNs emerged as an attempt at reducing the computational complexity of neural networks when applied to image recognition tasks [90]. Considering, for instance, a 128×128 pixel image with three channels for the RGB colours gives an input dimensionality of $128 \times 128 \times 3 = 49152$. Plugging this into the simple 4-layer MLP introduced in Section 5.1 gives a formidable $512 \cdot (49152 + 1) + 512 \cdot (512 + 1) + 10 \cdot (512 + 1) = 25434122$ weights. Larger pictures, and more sophisticated models, would increase this number polynomially, again presenting a familiar problem: the curse of dimensionality. Instead, the constraints of early hardware motivated the exploitation of correlations between local pixels. Features appearing in images are often local. A picture of a dog, for instance, will almost always contain two ears, but these will appear only in small corners of the image, rather than extending across it. Thus, the network should sample

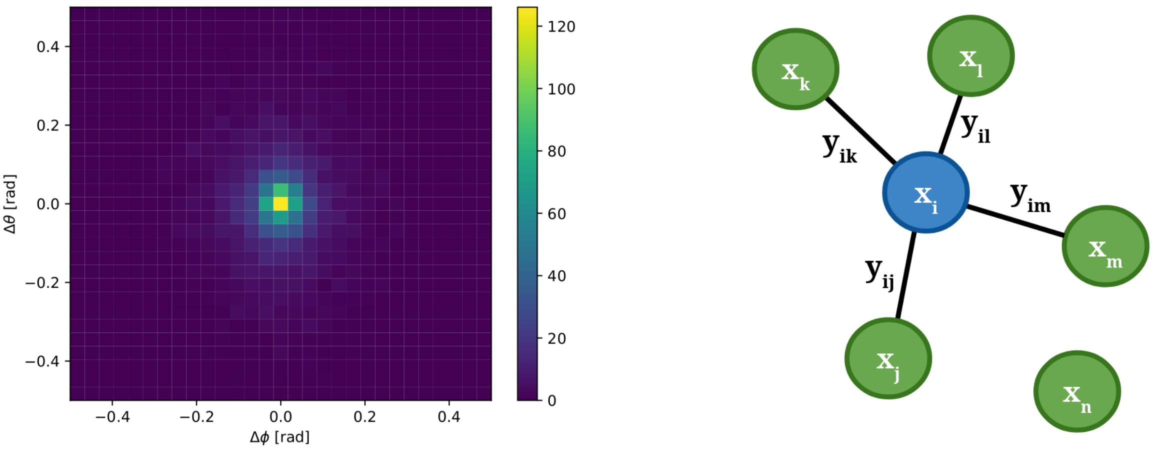


Figure 5.4: Two-dimensional $\theta - \phi$ distribution of K_L particles in strange jets from $e^+e^- \to Z \to q\bar{q}$ events at $\sqrt{s} = 91$ GeV. The image consists of 29×29 pixels and is centered at the jet axis. Pixel intensities are weighted by the $|p|/|p_{jet}|$ of the corresponding jet constituents (a). Simplified schematic of a graph representation of a jet. Nodes $\vec{x}_i$ represent jet constituents and their feature vectors, while edges $\vec{y}_{ij}$ encode differences between the features of connected nodes. Node $\vec{x}_i$ is connected only to its 4 nearest neighbours, highlighting some common limitations of GNNs (b).

small neighbourhoods of the image when aiming to extract features. Moreover, although the ears will be spatially separated, they anatomically share many similarities. A final challenge is that the dog would not always appear in the same position or orientation, and thus the network should be able to extract features regardless of where they appear in the image.

Convolutional neural networks address these challenges by introducing kernels (also called filters) that slide across the input and extract features by computing the cross-correlation between the input and the kernel, as depicted on Figure 5.5.

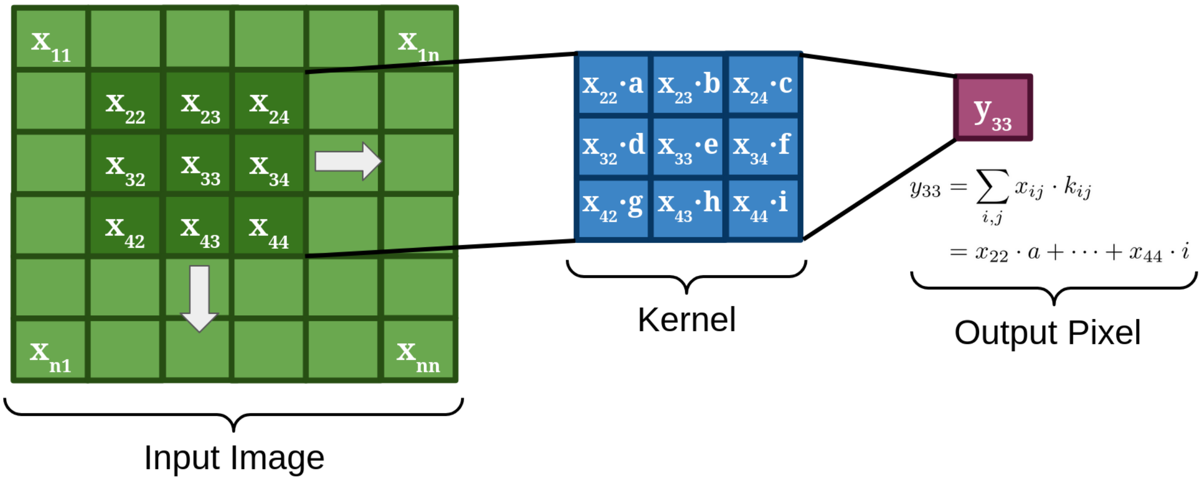

x_{11}
x_{1n}
x_{22}
x_{23}
x_{24}
x_{32}
x_{33}
x_{34}
x_{42}
x_{43}
x_{44}
x_{n1}
x_{nn}
Input Image
$x_{22}\cdot a$
$x_{23}\cdot b$
$x_{24}\cdot c$
$x_{32}\cdot d$
$x_{33}\cdot e$
$x_{34}\cdot f$
$x_{42}\cdot g$
$x_{43}\cdot h$
$x_{44}\cdot i$
Kernel
y_{33}
$y_{33} = \sum_{i,j} x_{ij} \cdot k_{ij}$
$= x_{22} \cdot a + \cdots + x_{44} \cdot i$
Output Pixel

Figure 5.5: Simplified schematic of a single output y_{33} of a convolutional layer consisting of a 3×3 kernel. Stride directions are indicated by the white arrows.

The output for each kernel $\mathbf{K}$ that is used on an $n \times n$ image $\mathbf{I}$ is an $m \times m$ image $\mathbf{S}$, where padding can be used to set $m = n$. Formally, a 2-d convolutional layer is defined by sliding the kernel along the input image and computing the Hadamard product between the kernel and the local elements of the input image and summing the output as

$$S_{i,j} = \sum_{c=1}^{C_{\text{in}}} \sum_{m=1}^{k_H} \sum_{n=1}^{k_W} I_{c,i+m,j+n} \cdot K_{c,m,n} \tag{5.29}$$

where C_{in} denotes the number of channels of the input image $\mathbf{I}$, and k_H, k_W denote the height and width of the kernel $\mathbf{K}$, respectively. The sliding of the kernel is controlled by the so-called *stride*, which describes for which elements $I_{i,j}$ the cross-correlation is computed. For instance a stride $s_H = 2$ along the height, and a stride $s_W = 2$ along the width, would entail the computation of the cross-correlation of every second element of $I_{i,j}$ along both dimensions, such that

$$i \in \{2k \mid k \in \mathbb{N},\ 2k \leq N_H\}, \quad j \in \{2l \mid l \in \mathbb{N},\ 2l \leq N_W\}$$

Computing the cross-correlation between the input and the kernels addresses several of the aforementioned challenges, since the kernels extract features of adjacent pixels, and are agnostic to where features appear in the image, since the kernel does not change as it is slid across the image. By introducing several, distinct kernels, whose outputs are passed to further convolutional layers, a variety of low-level features can be extracted by early filters which are combined into more sophisticated objects in later layers. Convolutional neural networks have enabled highly accurate image recognition, particularly following the introduction of GPU training [82, 91, 92]. In particle physics convolutional neural networks have been used to process detector images including calorimeter deposits as energy or p_T-weighted $\eta - \phi$ distributions of particles [93]. Despite the advantages with respect to the reduced number of weights that CNNs offer, they are not very well suited for sparse representations, where a large number of the pixels are equal to 0, such as those appearing in detector or particle images. Not only do sparse images entail computational inefficiency, but a main bottleneck in modern compute is the memory access speed, and thus modern jet tagging algorithms adopt more straightforward representations of jets, including the point cloud [94–96].

Recurrent Neural Networks

Recurrent neural networks (RNNs) first emerged in the late 80's and early 90's [97, 98], primarily in the context of natural language processing and time series modelling, as a way to address input data with temporal or sequential structure. Traditional approaches like the MLPs introduced in Section 5.1 assume independent inputs, and are thus not well suited for input data with sequential structure like natural language. While it is in principle possible to simply pass entire sequences to an MLP, this would entail using fixed-length sequences and increased complexity. While CNNs used kernels to exploit correlations between spatially adjacent inputs, RNNs preserve memory of previous inputs by introducing recurrent connections as a way to exploit correlations between temporally adjacent inputs. In this way weights are reused, and the symmetry of the problem is exploited as a way to reduce compute. Simple RNNs, or Elman networks [97], resemble 2-layer MLPs, but introduce a connection to the values of the hidden layer for the previous element of the input sequence. For the t^{th} element of some sequence $\vec{x}_1, ..., \vec{x}_n$, the output y_t is computed as

$$\vec{h}_t = g(\mathbf{U} \cdot \vec{h}_{t-1} + \mathbf{W} \cdot \vec{x}_t) \tag{5.30}$$

$$\vec{y}_t = f(\mathbf{V} \cdot \vec{h}_t) \tag{5.31}$$

where $\mathbf{U}$, $\mathbf{W}$, $\mathbf{V}$ are the weight matrices, and g, f are the activation functions. However, simple RNNs often struggle in propagating forward information that is distant in the input sequence. Moreover, the repeated multiplications that the hidden layers are subjected to in a recursive pass during the gradient computation can lead to vanishing and exploding gradients [99]. More sophisticated RNN architectures relying on gating mechanisms to control the flow of information, such as Long Short-Term Memory (LSTM) [100] and Gated Recurrent Unit (GRU) [101] networks, were introduced to address these challenges. RNNs have been used for jet tagging by representing jet constituents as elements of a sequence sorted according to their impact parameter or some other metric [102, 103], achieving considerably better performance than their MLP predecessors.

Graph Neural Networks

Graph neural networks (GNNs) are a class of neural networks that operate directly on a graph representation of the input data, consisting of nodes containing input feature vectors $\vec{x}_i$ and edges $\vec{e}_{ij}$ that encode relationships between node pairs. GNNs function by aggregating information of a node and its connected edges into an updated representation of the node in some latent space. The node is iteratively updated to encode information not only of itself, but also its relation to adjacent nodes, allowing for the extracted information to be processed by later layers in the neural network, in analogy to the feature extraction performed by successive convolutional layers in CNNs. GNNs are naturally permutation invariant unlike RNNs, making them an attractive candidate for tasks like jet flavour tagging, where the constituents of a jet do not display an inherent ordering.

GNNs have found a variety of applications in particle physics, including jet flavour tagging [94, 95, 104, 105], secondary vertex finding [105, 106], and particle reconstruction [107–109]. In the context of jet flavour tagging, jets are represented as sets of nodes, corresponding to jet constituents, and edges, corresponding to some pair-wise defined feature, often the difference in a given variable between jet constituents. In order to lower complexity, nodes of a graph are often only connected to a subset of its closest neighbours. Determining the k nearest neighbours, and according to what metric, is a non-trivial operation that cannot be neatly included in the backpropagtion given in Equation 5.25, as it is not differentiable. Moreover, long-range correlations between nodes that are distant in the graph may be neglected by considering only a subset of its neighbouring nodes. These limitations motivated the use of architectures which dynamically assigned weights to the connections between a node and all other nodes in a graph. The attention mechanism introduced in the following section addresses these challenges, and its adoption in jet flavour tagging resulted in architectures that were not only more performant, but also significantly faster to train [2, 110].

5.4 Transformers and the (Self-) Attention Mechanism

Attention first emerged in the context of machine translation [111], as a way to address challenges of pure RNN encoder-decoder architectures. Encoder-decoders [112, 113] function by transforming the input sequence $\{\vec{x}_1, ..., \vec{x}_n\}$ into some latent representation $\vec{c}$, which is then translated into an output sequence $\{\vec{y}_1, ..., \vec{y}_n\}$ by a decoder. While effective for short sequences, these architectures struggle for long inputs as the latent representation $\vec{c}$ is often taken to be the final hidden state of the encoder, and thus may not represent early elements of the input sequence well. Attention mechanisms were introduced as a way to dynamically assign weights to all elements of a sequence, thereby focusing on different parts of the sequence when relevant. In particular, rather than including only a single latent representation $\vec{c}$ of the input sequence, a set of representations $\{\vec{c}_i\}$ could be constructed by computing similarity scores between different elements of the sequence as:

$$\alpha_{ij} = \frac{\exp(\text{score}(\vec{h}^d_{i-1}, \vec{h}^e_j))}{\sum_l \exp(\text{score}(\vec{h}^d_{i-1}, \vec{h}^e_l))} \tag{5.32}$$

$$\vec{c}_i = \sum_j \alpha_{ij} \vec{h}^e_j \tag{5.33}$$

where the score() is some similarity score, α_{ij} are the attention scores, $\vec{h}^d_{i-1}$ is the decoder hidden state of the previous output sequence element, and $\vec{h}^e_j$ is the hidden state of the j^{th} element of the input sequence. More generally, attention mechanisms compute a score α_{ij} between a query vector $\vec{q}_i$ and a key $\vec{k}_j$ vector, which determines the contribution of a value vector $\vec{v}_j$ to the output. The key and the value vectors typically originate from the same set. The specific similarity score that is used to relate the query and the key defines the type of the attention mechanism, with variants including additive attention [114], dot-product attention [115], and its refinement, scaled dot-product attention [116]. Self-attention denotes the special case where the key, query, and value all originate from the same set.

Transformers are a class of neural networks that make use of a self-attention mechanism [116]. At its core, the self-attention mechanism captures the relations of an element of a sequence to all other elements of the same sequence. In this work, we will consider the self-attention mechanism in the form it was introduced in [116]: the scaled dot-product attention. Given an input sequence $\{\vec{x}_1, ..., \vec{x}_n\}$, each element is projected into a query[3] $\vec{q}_i = \vec{x}_i \mathbf{W}^Q$, key $\vec{k}_i = \vec{x}_i \mathbf{W}^K$, and value $\vec{v}_i = \vec{x}_i \mathbf{W}^V$ by the weight matrices $\mathbf{W}^Q$, $\mathbf{W}^K$, $\mathbf{W}^V$. The scaled dot-product attention is given as

$$\vec{y}_i = \sum_j \frac{\exp(\vec{q}_i \cdot \vec{k}_j / \sqrt{d_k})}{\sum_l \exp(\vec{q}_i \cdot \vec{k}_l / \sqrt{d_k})} \vec{v}_j = \sum_j \alpha_{ij} \vec{v}_j \tag{5.34}$$

where α_{ij} are attention weights, $\vec{y}_i$ are the output vectors, and d_k is the dimension of $\vec{q}$, $\vec{k}$. In summary, the output of the self-attention mechanism is the sum of the values $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ scaled by the attention weights α_{ij}, which reflect the importance of the j^{th} input to the i^{th} output vector via the softmaxed dot product of the query $\vec{q}_i$ and key $\vec{k}_j$.

One of the key motivations for Transformer NNs is that the computation of $\vec{y}_i$ in equation 5.34 does not depend on any other $\vec{y}_{k \neq i}$, i.e. there is no recursion. Moreover, equation 5.34 can be rewritten in matrix form by defining $\mathbf{Q} = (\vec{q}_1^\intercal | ... | \vec{q}_n^\intercal)^\intercal$, $\mathbf{K} = (\vec{k}_1^\intercal | ... | \vec{k}_n^\intercal)^\intercal$, $\mathbf{V} = (\vec{v}_1^\intercal | ... | \vec{v}_n^\intercal)^\intercal$. The scaled dot-product attention $\mathbf{Y} = (\vec{y}_1^\intercal | ... | \vec{y}_n^\intercal)^\intercal$ is then given as

$$\mathbf{Y} = \text{softmax}(\frac{\mathbf{Q}\mathbf{K}^\intercal}{\sqrt{d_k}})\mathbf{V} \tag{5.35}$$

---

[3] Throughout this section a row-vector convention is adopted, in accordance with standard treatments of transformer models. This departs from the column-vector convention adopted for previous sections in this chapter.

Complex dependencies between elements are not fully captured by a single attention weight. Transformers employ multiple *attention heads* in parallel, by defining a set of weight matrices for the queries $\{\mathbf{W}_l^Q\}$, keys $\{\mathbf{W}_l^K\}$, and values $\{\mathbf{W}_l^V\}$. The dimensionality of the query and key vectors is typically set to the ratio of the input dimensionality and the number of heads, $d_k = d/n_h$, such that each attention head operates in a d_k-dimensional subspace of the input. The scaled dot-product attention is then computed for each set l of queries, keys, and values according to Equation 5.35, resulting in a set of output scores where $\mathbf{h}_l = \mathbf{Y}_l \in \mathbb{R}^{n\times d_v}$, where d_v denotes the dimension of the value vectors. The output of all n_h attention heads is concatenated along the feature dimension, resulting in a matrix $\mathbf{H} \in \mathbb{R}^{n\times(n_h\cdot d_v)}$, which is projected back into $\mathbb{R}^{n\times d}$ by a weight matrix $W^O \in \mathbb{R}^{(n_h\cdot d_v)\times d}$:

$$\mathbf{h}_l = \text{Attention}(\mathbf{Q}^l, \mathbf{K}^l, \mathbf{V}^l), \quad \mathbf{H} = (\mathbf{h}_1|...|\mathbf{h}_{n_h})$$

$$\text{MultiHeadAttention}(\{\vec{x}_1, ..., \vec{x}_n\}) = \mathbf{H}\cdot\mathbf{W}^O \tag{5.36}$$

Multi-head attention (MHA) is the workhorse of Transformer models. Figure 5.6 depicts a schematic representation of multi-head attention on an input set $\{\vec{x}_1, ..., \vec{x}_n\}$.

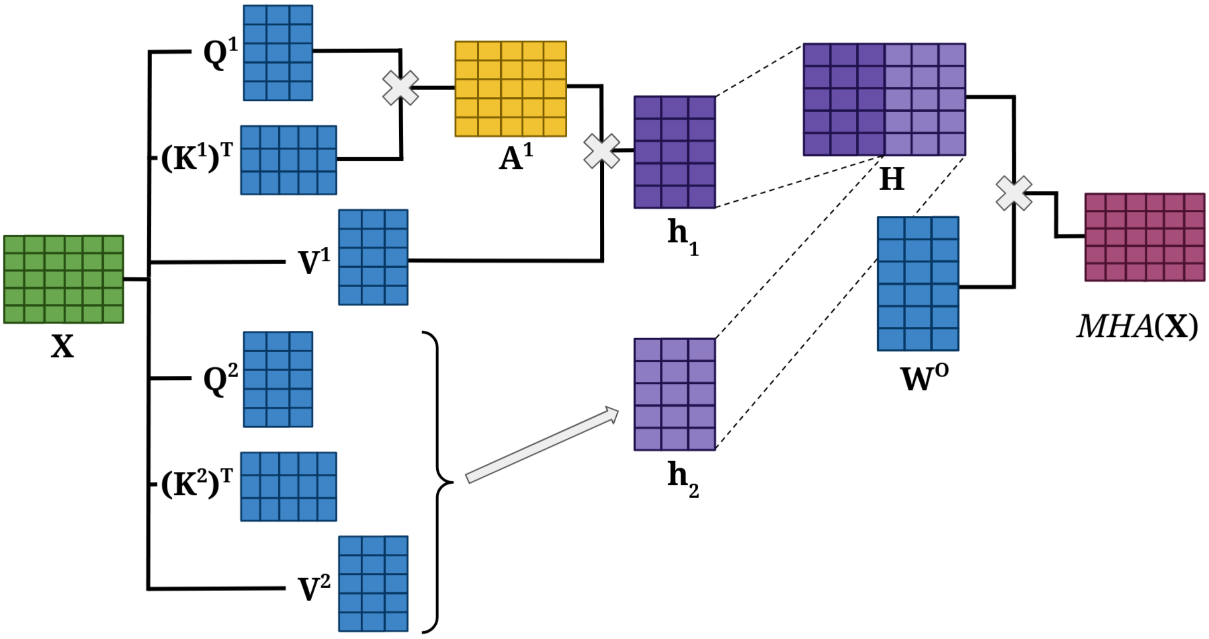


Figure 5.6: Simplified schematic of a 2-head MHA layer on input sequence $\mathbf{X} = (\vec{x}_1^\intercal|...|\vec{x}_n^\intercal)^\intercal$. The white × denotes matrix multiplication. Query, key, and value projections, as well as softmax and $1/\sqrt{d_k}$ scaling, are omitted for clarity. Blue matrices denote learnable parameters.

The fact that the computation of self-attention can be written as matrix multiplications, according to Equation 5.35, highlights one of the key advantages of Transformers with respect to previous architectures: parallelisability. In practice, this means that Transformers can often be trained and perform inference considerably faster than competing architectures such as RNNs or GNNs.

Given the fact that Transformers capture correlations between potentially distant neighbours (in the case of jet tagging, these could be jet constituents) and also come with sizeable computational advantages, several applications already exist in HEP [2, 96, 110, 117].

In the context of FCC-ee feasibility studies, fast training is of utmost importance, as present

NN architectures are being used to gauge the physics potential of different detector configurations, to each of which a NN will have to be optimised. The Z boson run will be the most challenging computationally with roughly 4.8×10^{12} visible decays [29]. Moreover, given the 20+ year timeline, until the FCC-ee would be in operation, it is certain that current ML approaches will be superseded by more performant architectures.

5.5 Machine Learning in Jet Flavour Tagging

Machine learning techniques are uniquely suited to the task of discriminating jet flavours, where training samples are abundant in the form of MC simulation, but the underlying dynamics are poorly understood. Jet flavour tagging is often sub-divided into flavour tagging, which aims to identify b, c, s, and light quarks, and jet substructure tagging, which targets W, H, Z, t, and gluons. A primary distinction is how the discriminating information manifests [93]. In the case of (heavy) flavour tagging, the long lifetime of B/D-mesons results in displaced vertices. By contrast, gluon jets do not display such identifiable structure, and much of the distinguishing information is encoded in the distributions of the outgoing hadrons. This section provides a brief overview of machine learning in flavour tagging, mostly from an LHC perspective. The interested reader is encouraged to consult the comprehensive review of jet flavour tagging at the LHC found in Ref. [69].

Clearly, there were many ML developments in jet tagging prior to the LHC, but the LHC-era has coincided with the shift to the GPU paradigm, and thus modern machine learning. Early applications of ML to jet tagging revolved primarily around simple MLPs trained on hand-engineered features and date back to LEP [118–121].

The introduction of BDTs at the MiniBooNE experiment [122, 123] and at the D0 experiment [124, 125], shaped early ML architectures at the LHC, particularly during Run 2. The computational advantages of BDTs were leveraged by providing algorithms with carefully curated high-level variables that had been shown to be discriminating between jet flavours, as those presented in Section 6.2. BDTs often outperformed their simple MLP counterparts [126].

Initial jet tagging algorithms at the CMS and ATLAS experiments relied on likelihood ratios constructed from high-level physics-motivated variables [127–129] Improvements were observed with the inclusion of MLPs, such as from the Combined Secondary Vertex (CSV) algorithm [127] to its successor, CSVv2 [55, 130]. Simple MLP architectures were used to combine the outputs of several individual tagging algorithms, both at CMS [55, 130] and at ATLAS [128, 131]. With Run 2 of the LHC came the advent of BDTs in the form of cMVAv2 [55] at CMS and MV2 [132, 133] at ATLAS, leading to sizeable improvements in performance [69].

BDT-based architectures were eventually superseded by MLPs in jet tagging as the shift to GPU-training allowed models to become deeper and include more input features. Dense Neural Networks (DNNs), that is to say fully-connected neural networks with several hidden layers, have been used in the context of jet tagging using a variety of inputs. Ref. [134], for instance, uses physics motivated high-level variables as inputs for quark gluon discrimination. The DeepCSV algorithm [55, 135] incorporated similar features to its predecessor, but with an increased number of tracks and model complexity leading to a performance increase. The DL1 algorithm [136, 137] similarly implemented a DNN, with many of the same features as MV2.

The emergence of CNNs occurred mostly in the realm of jet substructure tagging. Contrary to the secondary vertex variables that proved discriminating for heavy flavour jets, quark gluon tagging relies on variables that capture the spatial distribution and multiplicity of outgoing partons, which hint at the colour factor of the initiating parton. Motivated by the success of CNNs in computer vision, CNNs were applied to quark-gluon discrimination [138–140] and top tagging [141].

RNNs allowed the processing of more complex dependencies between jet constituents. DeepJet [102], in particular, is tailored to processing large input sequences, and allowed the processing of considerably more information, including that of neutral jet constituents. DeepJet achieved a 20% improvement with respect to DeepCSV in b tagging efficiency at a light-jet mistag rate of 0.1% [69].

Graph representations of jets [94, 95, 104, 105] emerged as one of the most performant representations. The ParticleNet [94] algorithm uses EdgeConvolutions [142] in order to extract correlations between jet constituents, outperforming previous CNN architectures in quark-gluon and top tagging [94], and RNN architectures in heavy flavour tagging [69].

With the introduction of Transformers [2, 96, 110, 117], jet tagging algorithms addressed two shortcomings of pure GNNs: computational efficiency and full connectivity among elements of an input sequence. ParT [110] and GN2 [117] achieve similar, or improved, performance with respect to their pure GNN counterparts [69]. Transformers have emerged as a considerably more efficient architecture, with ParT displaying a reduced number of FLOPs, despite including vastly more parameters [110].

Recent developments in jet flavour tagging have been marked by a general trend to process low-level inputs with ever more complex models. While the additional information, coupled with the additional expressivity of the models, has resulted in demonstrable improvements in tagging efficiencies, a concern has become amplified: the task $\mathbf{T}$ does not perfectly align with the discrimination of actual jets. Indeed, while jet tagging algorithms are primarily trained on labelled simulations, the ultimate goal is to perform inference on detector data. Sophisticated models are particularly susceptible to learning unphysical artifacts of the simulation as distinguishing features. For instance, the classification of quark and gluon jets has been observed to perform differently with different generators such as Pythia and Herwig, with quark and gluon jets appearing more distinct in Pythia than in Herwig [138]. Moreover, classification performance is also sensitive to the particular jet flavour definition that is adopted, including those introduced in Section 4.3. Much emphasis is put on the Data/MC agreement and the computation of scale factors in an effort to quantify the errors carried over from simulation. Promising avenues to better align the task $\mathbf{T}$ with the discrimination of true jets, rather than merely optimising a pure data science classification problem on simulations, include adversarial [143, 144], as well as weakly- and unsupervised [145–149] learning techniques. Ref. [93] expands on this discussion, and highlights some further problems that emerge from the trade-off of performance against systematic uncertainties and interpretability.

6 Jet Flavour Tagging at the FCC-ee with a Transformer-based Neural Network

The identification of jets is instrumental to the exploitation of the full potential of collider experiments. The steady evolution of the machine learning techniques presented in the previous chapter has led to ever increasing accuracy in the task of jet flavour tagging. Nevertheless, the relatively messy environment at hadron colliders has meant that jet flavour tagging efforts have focused primarily on heavier b- and c-jets, whose distinctive signatures allows them to be disentangled from the formidable QCD background. The FCC-ee, and lepton colliders in general, offer much cleaner environments in which the initial state kinematics are known, without the additional complications of parton distribution functions or initial state QCD radiation. The prospect of tagging strange jets at the FCC-ee would enable studies of processes such as $Z \to s\bar{s}$ production, rare Higgs decays to s-jets, CKM matrix elements via W decays, and BSM physics scenarios such as FCNCs at tree level.

DeepJetTransformer [1, 2], a multiclassifier neural network using a transformer-based architecture, is coupled with secondary vertexing and K-short reconstruction at the FCC-ee to discriminate strange quark initiated jets. This chapter begins by introducing the input features and network architecture of DeepJetTransformer, followed by details of the training procedure. The classification results of DeepJetTransformer on $Z \to q\bar{q}$ jets at $\sqrt{s} = 91.2$ GeV and $Z(\to \nu\nu)H(\to q\bar{q})$ jets at $\sqrt{s} = 240$ GeV are presented, with a focus on strange jet discrimination. Subsequently, the importance of individual features to the classification performance is evaluated. Finally, the chapter concludes with a case study demonstrating the isolation of $Z \to s\bar{s}$ jets using DeepJetTransformer, before the main findings are summarised.

> *The jet flavour tagging results covered in this chapter have been published in Ref. [2]. Parts of the text have been reproduced from the original publication.*

6.1 Jets

Jets were clustered from $Z \to q\bar{q}$ events at $\sqrt{s} = 91.2$ GeV and $Z(\to \nu\nu)H(\to q\bar{q})$ events at $\sqrt{s} = 240$ GeV. The events were generated using Pythia 8.303 [150], including parton showering and hadronisation, while the detector response was simulated using Delphes [151], assuming the IDEA [152, 153] detector concept described in Section 3.5. Charged tracks were reconstructed with efficiencies of 99.7%, 65%, and 4% in the high (> 0.5 GeV), medium (0.5-0.3 GeV), and low (< 0.3 GeV) momentum ranges, respectively. No fake tracks were included in the simulation.

The exclusive e^+e^- k_T algorithm [64], implemented in FastJet 3.3.4 [63], was used for clustering the particle flow style objects output by Delphes. While other algorithms, including the anti-k_T algorithm [62] and the generalised e^+e^- k_T [63], were explored, the e^+e^- k_T algorithm was found to be the most suitable due to the simple event topology, and its robustness to gluon emissions (i.e. $Z \to q\bar{q}g$).

Jet flavour labels were defined as the flavour of the outgoing partons from the boson decay. While this provides a natural and unambiguous definition in the 2-parton decays studied (e.g. $Z \to q\bar{q}$), it does not generalise to multi-parton decays such at $ZH \to 4$ jets, or allow the sub-division into quark and antiquark jets. It is worth noting that this jet labelling approach departs from the ghost-matching [66] technique commonly used at hadron colliders, where jet flavour is defined by the reclustering of particles with partons from earlier stages in the Monte Carlo history.

6.2 Input Features

The input to DeepJetTransformer consists of 5 distinct classes: jet-level variables, charged jet constituent variables, neutral jet constituent variables, secondary vertex variables, and V^0 variables. The first of these, jet-level variables, are either kinematic variables derived from the 4-vector of the jet, or properties of the entire jet derived from its constituents. These include variables such as the number of charged constituents or the various jet angularities constructed from the momentum-weighted spatial distributions of the jet constituents. The full set of jet-level variables used in DeepJetTransformer is listed in Table 6.1.

Table 6.1: Summary of per-jet global input variables to DeepJetTransformer.

Input Feature	Description
$\lvert p\rvert,\ E,\ m$	3-momentum magnitude, energy, and invariant mass of the jet
θ, ϕ	polar and azimuthal angle of the jet axis
$\mathrm{N}_{\mathrm{charged}}$	charged particle (track) multiplicity in the jet
$\mathrm{N}_{\mathrm{neutral}}$	neutral particle multiplicity in the jet
$\lambda^{\kappa}_{\beta} = \Sigma_{i \in \mathrm{jet}} z_i^{\kappa} R_i^{\beta}$	jet angularity [154] as sum of normalized jet constituent energy (z_i) and angular distance to jet axis (R_i) for ($\kappa = 0$, $\beta = 0$), ($\kappa = 1$, $\beta = 0.5$), ($\kappa = 1$, $\beta = 1$), ($\kappa = 1$, $\beta = 2$), ($\kappa = 0$, $\beta = 2$)

Jet constituents themselves are split into charged particles and neutral particles, corresponding to the sub-detectors that were used for their reconstruction in the particle flow-like algorithm [155]. For both charged and neutral particles, a set of kinematic variables related angle and momentum with respect to the jet (axis) is defined, as listed on Table 6.2. For each jet up to 25 charged jet constituents and 25 neutral jet constituents are considered. This is enforced by truncating the input feature array of a given jet if the number of charged/neutral jet constituents is more than 25. Conversely, if the number of charged/neutral jet constituents is less than 25, then the input feature array is zero-padded.

Charged particles additionally include track-specific variables such as the signed transverse and longitudinal impact parameters, which describe the distance of closest approach to the primary vertex for a given track. Figure 6.1a depicts the transverse impact parameter significance, which serves as a distinguishing feature as tracks with a large transverse impact parameter originate overwhelmingly from heavy jets. Finally, a set of flags identifying the particle type is included for charged and neutral particles. Many of these flags are naturally output by particle flow-like algorithms with high precision. Thus in this implementation, they are perfectly reconstructed from the MC truth. This includes the "isMuon" and "isElectron" flags for charged particles, as well as the "isPhoton" flag for neutral particles, which tellingly describe the type of a given jet constituent. The "isKaon" flag identifies a particle as a charged Kaon ($K^{\pm}$) and

Table 6.2: Summary of jet constituent input features, divided into charged and neutral jet constituents.

Input Feature	Description
$D_0(z_0)$	signed transverse (longitudinal) impact parameter
$D_0/\sigma_{D_0}(z_0/\sigma_{z_0})$	signed transverse (longitudinal) impact parameter significance
$\theta_{\text{rel}}(\phi_{\text{rel}})$	polar (azimuthal) angle of track with respect to the jet axis
R	angular distance of track and jet axis
C	half-curvature of the track
$m_{\text{ch.}}, q$	track invariant mass and charge
$\frac{\lvert p\rvert_{\text{ch.}}}{\lvert p\rvert_{\text{jet}}}, \ln(\lvert p\rvert_{\text{ch.}}), \ln\left(\frac{\lvert p\rvert_{\text{ch.}}}{\lvert p\rvert_{\text{jet}}}\right)$	(normalised) magnitude of track momentum and logarithms
$\frac{E_{\text{ch.}}}{E_{\text{jet}}}, \ln(E_{\text{ch.}}), \ln\left(\frac{E_{\text{ch.}}}{E_{\text{jet}}}\right)$	(normalised) track energy and logarithms
`isKaon`	if the particle is identified as a $K^{\pm}$
`isMuon`	if the particle is identified as a $\mu^{\pm}$
`isElectron`	if the particle is identified as an $e^{\pm}$
$\theta_{\text{rel}}(\phi_{\text{rel}})$	polar (azimuthal) angle of particle with respect to the jet axis
R	angular distance of neutral particle and jet axis
$\frac{\lvert p\rvert_{\text{neut.}}}{\lvert p\rvert_{\text{jet}}}, \ln(\lvert p\rvert_{\text{neut.}}), \ln\left(\frac{\lvert p\rvert_{\text{neut.}}}{\lvert p\rvert_{\text{jet}}}\right)$	(normalised) magnitude of particle momentum and logarithms
$\frac{E_{\text{neut.}}}{E_{\text{jet}}}, \ln(E_{\text{neut.}}), \ln\left(\frac{E_{\text{neut.}}}{E_{\text{jet}}}\right)$	(normalised) neutral particle energy and logarithms
`isPhoton`	if the particle is identified as a Photon

is thus central to the identification of strange jets. Figure 6.2a depicts the charged Kaon multiplicity of s-, u-, and d-jets, where it can be seen that Kaons are more prevalent in strange jets. This effect is amplified when particle momentum is accounted for, as discussed in Ref. [156]. The identification of charged Kaons relies on distinguishing Kaons from other charged hadrons, primarily pions and protons. Cluster counting and time-of-flight have emerged as promising particle identification techniques, with separations of $> 3\sigma$ achievable in the momentum range of $|p| < 30$ GeV with the IDEA detector concept [59]. In this work the identification of charged Kaons is emulated by sampling a uniform distribution with a $K^{\pm}$ efficiency of $\epsilon_{sig}^{K^{\pm}} = 90\%$, and a background efficiency of $\epsilon_{sig}^{\pi^{\pm}} = 10\%$, with charged pions serving as the false positive for the "isKaon" flag. This constitutes the *baseline* scenario of DeepJetTransformer, although some additional scenarios are explored in Section 6.5.1. The deliberately conservative estimate reflects PID studies at Belle, where an efficiency of $\epsilon_{sig}^{K^{\pm}} = 87.99 \pm 0.12\%$ with a fake rate of $\epsilon_{sig}^{\pi^{\pm}} = 8.53 \pm 0.10\%$ was achieved for particles below 4 GeV [157].

The final two classes of features are SVs and V^0s, both of which originate from the vertexing algorithm acting on charged particle tracks. The set of SV and V^0 features is listed on Table 6.3. The V^0 and SV input feature arrays are likewise truncated/zero-padded, with an array length of 4 being enforced. The distinguishing power of some of these variables is discussed below.

The reconstruction of SVs has long been the cornerstone of heavy jet identification, and begins with the identification of tracks originating from the primary vertex. Figure 6.1b depicts the number of secondary vertices reconstructed for b-, c-, and light-jets, illustrating that jets containing at least 1 secondary vertex originate almost exclusively from heavy jets. As the primary goal of vertexing algorithms is the reconstruction of secondary vertices associated with the decay of b or c quarks, neutral particles decaying to a pair of charged tracks, denoted V^0s, are often reconstructed, and subsequently discarded. This is done in order to prevent the erroneous

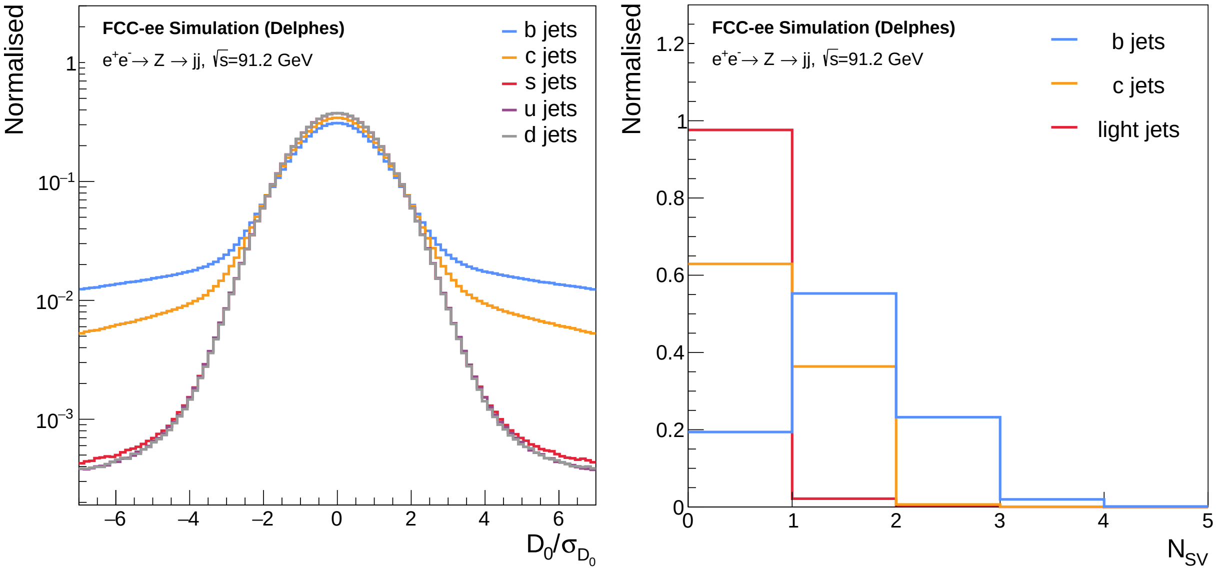


Figure 6.1: Distributions of the transverse impact parameter significance (a) and the number of secondary vertices (b) in $Z \to q\bar{q}$ jets at $\sqrt{s} = 91.2$ GeV. The distributions are split by MC flavour, highlighting their discriminating power in heavy-flavour tagging.

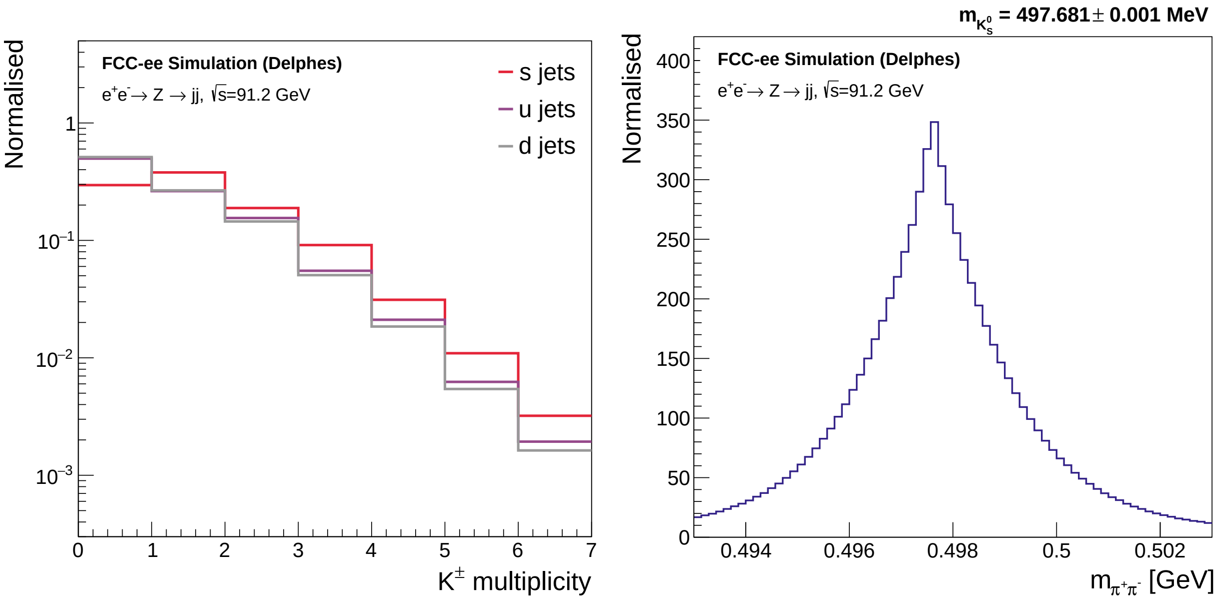


Figure 6.2: Distributions of the $K^\pm$ multiplicity (a) and track-pair invariant mass under the $K^0_S \to \pi^+\pi^-$ mass hypothesis (b) in $Z \to q\bar{q}$ jets at $\sqrt{s} = 91.2$ GeV.

Table 6.3: Summary of input features characterising each reconstructed secondary vertex. Input features are defined analogously for V^{0}s, with the addition of the PDG ID [158].

Input Feature	Description
$\lvert p \rvert,\ m$	3-momentum magnitude and invariant mass of the SV
$\mathrm{N}_{\mathrm{tracks}}$	track multiplicity of the SV
$\chi^2, \mathrm{N}_{\mathrm{DoF}}$	χ^2 and number of degrees of freedom of the SV
$\theta_{\mathrm{rel}}, \phi_{\mathrm{rel}}$	polar and azimuthal angle of the SV with respect to the jet axis
$\hat{\mathrm{p}}.\hat{\mathrm{r}}$	collinearity of SV with respect to PV
$\mathrm{d}_{\mathrm{3D}}, \mathrm{d}_{xy}$	3D and transverse distance of the SV from the PV

association of these decaying neutrals with bottom or charmed hadrons. Subsequently, remaining tracks undergo a fitting procedure to reconstruct secondary vertices, known as vertexing. In this work, an implementation of the vertexing module of the LCFIPlus framework [56, 159] in FCCAnalyses [160] was used, using a χ^2 vertex fitter [161]. One of the key differences in this implementation with respect to other vertexing approaches is that rather than reconstructing V^{0}s only for the purpose of discarding their tracks, V^{0}s are saved as separate objects within the jet. The finding of V^{0}s begins with the assumption of a mass hypothesis related to the V^0 decay under consideration. The decays $K_S^0 \to \pi^+\pi^-$ and $\Lambda^0 \to p\pi^-$ were included in this study. For instance, assuming a mass hypothesis of the Λ^0 would entail assigning one of the tracks the mass of a proton and the other the mass of a charged pion. Using the measured 3-momentum, the invariant mass of the track pair is computed. Figure 6.2b depicts the invariant mass of track pairs under the K_S^0 mass hypothesis. The pronounced peak at approximately the mass of the K_S^0 confirms the efficacy of the V^0 reconstruction algorithm. Track pairs are subsequently subjected to a set of constraints, in order to determine whether or not they originate from a V^0 decay. These revolve around how close the reconstructed invariant mass is to the true mass of the V^0 particle under consideration, and whether or not the V^0 is displaced enough from the primary vertex. Additionally, the track pair 3-momentum must be collinear with the vector linking the primary vertex to the V^0. The exact constraints that were adopted can be found in Ref. [2].

6.3 Network Architecture

The architecture of DeepJetTransformer, depicted on Figure 6.3, can largely be split into three sections: the convolutional layers that encode the input, the Heavy-Flavour (HF) transformer blocks, and the fully connected layers at the end of the network. The input to DeepJetTransformer consists of 5 classes of variables detailed in Section 6.2. All variable classes barring jet-level variables are passed through 1-d convolutional layers with 1 × 1 filters and stride 1. In the context of convolutional layers, introduced in Section 5.3, the 1 × 1 filters act as linear projections of each input constituent. By using the same filters for the entire input set, weights are shared for each of the constituents. This is desirable physically, as all constituents of a given class (e.g. charged particles) should be subjected to the same embedding. Moreover, it is worth noting that this is a suitable encoding for jet constituent variables, since the length of the input (e.g. number of charged particle variables) is not constant. More specifically, and focusing only on charged particles (though valid for all other classes), denote by $\vec{v}_{ch.}$ the m-dimensional vector of input features for a single charged particle in a jet. Then for a jet with charged multiplicity n, the charged particles variables will be a set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. The 1-d convolution operation encodes the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ into a new set of equal length $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ by defining the components of $\vec{u}_i$ as

$$(\vec{u}_i)_k = f(\vec{v}_i \cdot \vec{w}_k) \tag{6.1}$$

where $(\vec{u}_i)_k$ denotes the k^{th} component of the encoded vector $\vec{u}_i$, f is the ReLU activation function, and the vectors $\vec{w}_k$ are weight vectors corresponding to the filters of the convolution. The number of filters l determines the dimension of each output vector $\vec{u}$. Clearly, the 1-d convolution operation could be applied again on the set $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ with a new set of l' filters to obtain a further encoded set $\{\vec{u}'_1, \vec{u}'_2, \ldots, \vec{u}'_n\}$. In DeepJetTransformer the first 1-d convolution encodes the input $\{\vec{v}\}$ of n charged particles with 19 features (e.g. constituent $|p|$) into a set of n 64-dimensional vectors $\{\vec{u}\}$. These are then passed to a second 1-d convolution, which in addition introduces skip-connections $\vec{s}_i$ that are added to Equation 6.1, such that $(\vec{u}_i)_k \rightarrow (\vec{u}_i)_k + (\vec{s}_i)_k$. The skip connections are typically included in order to improve gradient flow, but in the case of the second convolution are defined by a separate 1-d convolution acting on $\{\vec{u}\}$, as in Equation 6.1, resulting in a set of n 128-dimensional vectors $\{\vec{u}^s\}$. Effectively, the set $\{\vec{u}\}$ is encoded into a set of n 128-dimensional vectors $\{\vec{u}'\}$, by adding the output of two separate 1-d convolutions. These are finally passed to a third 1-d convolution that encodes $\{\vec{u}'\}$ into a new set of n 128-dimensional vectors $\{\vec{u}''\}$, again employing skip-connections, where the components of the skip vector are now given by the input set itself: $\vec{s}_i = \vec{u}'_i$. The set $\{\vec{u}''\}$ is concatenated with the sets resulting from the encoding of other variable classes: $\{\vec{u}''_{neut.}\}$, $\{\vec{u}''_{SVs}\}$, $\{\vec{u}''_{V0_s}\}$, which are constructed in the same way. It is worth noting that while the 1-d convolution may alter the dimension of the vectors in the set it acts upon, it does not alter the length of the set. Each 128-dimensional vector $\vec{u}''$ corresponds to a single charged particle in the jet. In particular, the 1-d convolution does not capture any contextual information of the charged particle with respect to its neighbours. This instead happens in the next section of the network: the HF transformer blocks.

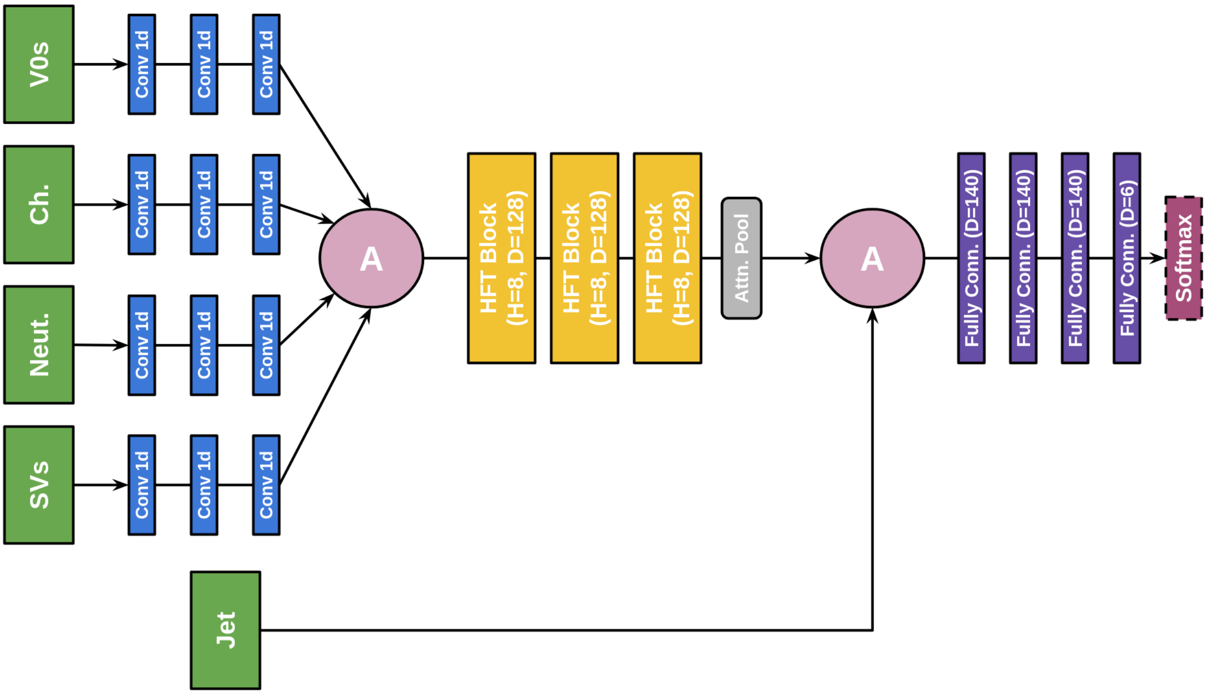


Figure 6.3: Schematic overview of the DeepJetTransformer architecture, illustrating the processing of input from left to right. Convolutional layers are greatly simplified, with details like the skip-connections described in Section 6.3 omitted for clarity.

In the first step of the HF transformer blocks, the output of the convolutional layers above $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\} = \{\vec{u}''_{ch.}\} \oplus \{\vec{u}''_{neut.}\} \oplus \{\vec{u}''_{SVs}\} \oplus \{\vec{u}''_{V^0s}\}$ is encoded via a 1-d convolution with 128 filters, with a skip-connection again defined as the input $\vec{s_i} = \vec{x}_i$. Next, the encoded set $\{\vec{x}'\}$ is used to define a set of queries $\mathbf{Q}$, keys $\mathbf{K}$, and values $\mathbf{V}$, in order to compute the scaled dot-product attention, as defined in equation 5.35. This is repeated $n = 8$ times for different weight matrices W_i^Q, W_i^K, W_i^V, and concatenated before being projected back to m 128-dimensional vectors $\{\vec{y}_{MHA}\}$, as defined in Equation 5.36. Subsequently, the output of the Multi-head attention step is added to the input via a skip connection such that $\vec{y_i} = \vec{y}_i^{MHA} + \vec{x}'_i$, before finally being passed to a pair of fully connected layers. The first fully connected layer projects $\{\vec{y}\}$ to a set of 512-dimensional vectors $\{\vec{y}'\}$ followed by a ReLU activation function, whereas the second serves simply to linearly project the set $\{\vec{y}'\}$ back to m 128-dimensional vectors $\{\vec{y}''\}$. Finally, a skip-connection is employed to define the output of the HF transformer layer as $\vec{y}_i^{out} = \vec{y}''_i + \vec{y}_i$. The output set $\{\vec{y}_{out}\}$ serves as input to another HF transformer block. This is stacked 3 times in DeepJetTransformer, enabling the attention mechanism to act on progressively more abstracted features.

The final section of DeepJetTransformer are the fully connected layers, where higher level features output by the HF transformer blocks are abstracted into class scores for each jet-flavour. In particular, the output of the HF transformer blocks $\{\vec{y}_{out}\}$ is passed through an attention pooling layer. The pooling layer begins by computing a score for each element in the sequence as a linear combination of all 128 features of each element, via a 1-d convolution with a single filter followed by a softmax activation function. This reduces Equation 6.1 to

$$\alpha_i = f(\vec{y}_i^{out} \cdot \vec{w}) \tag{6.2}$$

where α_i are scalar scores for each element of $\{\vec{y}_{out}\}$ that will be used for the pooling, $\vec{w}$ is the filter of the convolution, and f denotes the softmax function applied over all i. The elements of $\{\vec{y}_{out}\}$ are then pooled into a single output vector

$$\vec{y}_{pool} = f\left(\sum \alpha_i \cdot \vec{y}_i^{out}\right) \tag{6.3}$$

where f denotes a ReLU activation function applied element-wise. The vector $\vec{y}_{pool}$ is subsequently concatenated with the jet-level variables, before being passed to the fully connected layers. There are four stacked fully connected layers consisting of $140, 140, 140$, and 6 nodes, respectively, which were found to be performant hyperparameter choices. The first three layers are followed by a ReLU activation function, and a skip connection that adds the input of the layer to the output. The final layer simply projects the vector to a vector of dimension c, corresponding to the number of jet flavours. During inference, the raw network scores are normalised by a softmax function, given in Equation 5.7. These normalised outputs, which are constrained to lie between 0 and 1 and sum to 1 across all jet flavours, are referred to as "softmaxed classifier outputs" for brevity.

Throughout this discussion some details, including normalisation and dropout layers, have been omitted for clarity. Batch normalisation is used throughout the network, particularly following convolutional and linear layers, while layer normalisation is used in the HF Transformer blocks to improve training stability. The full implementation, along with documentation for the sample preparation, the training methodology described in the following section, and the relevant code, is publicly available here: DeepJetFCC[1]

---

[1] https://github.com/Edler1/DeepJetFCC/tree/master/docs

6.4 Training Details

Training and inference of the DeepJetTransformer algorithm was performed in Pytorch (v1.10.1) [85]. As detailed in Section 5.2, the training of a neural network entails the computation of weights that minimise the loss function, which in this case was chosen to be the categorical cross-entropy $L(\vec{t}, \vec{y})$, defined in Equation 5.11. By punishing incorrect predictions of each node separately, a multi-class classifier was trained for which each output node corresponds to a distinct jet flavour: b, c, s, u, d, (g). Weights were optimised by combining the rectified Adam optimiser [88], given in Equation 5.23, with the Lookahead optimiser [162]. The Lookahead optimiser introduces a set of "slow" weights alongside a set of "fast" weights computed by the RAdam optimiser. After k updates of the fast weights, the slow weights are updated by moving a fraction α towards the fast weights, improving training stability, particularly for aggressive weight updates. During training, the number of updates was set to $k = 6$, while the fraction was set to $\alpha = 0.5$. An initial learning rate η of $5 \cdot 10^{-3}$ and decay rates β_1, β_2 of 0.95, 0.999 were chosen for the RAdam optimiser.

The training dataset is comprised of roughly 1 million jets from $Z \to q\bar{q}$ events at $\sqrt{s} = 91.2$ GeV, split roughly evenly among the different jet flavours. The dataset was divided 80/20% into training and validation datasets. Training was conducted for 70 epochs with the learning rate being unchanged for the first 49 epochs, but decaying exponentially thereafter. A relatively large batch size of 4000 was used for training stability. Figure 6.4 depicts the training and validation loss for each epoch during the training of the baseline version of the classifier. Upon inspection, an epoch at which the training training and validation losses were similar was selected, before the onset of overfitting and subsequent divergence of the loss. With about 10^6 parameters and efficient transformer blocks as the workhorse, training DeepJetTransformer takes about 2 hours on an NVIDIA Tesla V100s GPU. Finally, the model's performance was evaluated on a separate set of 1 million jets, with similar balancing of jet flavours, to which it was not exposed during training.

The tagger was trained separately on $Z(\to \nu\nu)H(\to q\bar{q})$ events at $\sqrt{s} = 240$ GeV for the $H(\to q\bar{q})$ results shown in Section 6.5. The focus of these studies was jet flavour tagging at the Z resonance, with the classification of $H \to q\bar{q}$ events serving primarily as a comparison to the classification performance of other jet flavour taggers for future colliders, like ParticleNetIDEA [59, 163].

6.5 Classification Results

A binary classifier was constructed for each jet flavour $q \equiv u, d, s, c, b, (g)$ with a signal flavour (i) and a background flavour (j):

$$S_{ij} = \frac{S_i}{S_i + S_j}, \tag{6.4}$$

where S_i are the softmaxed classifier outputs, shown in Figure 6.5 for the five quark flavours.

ROC curves were computed for each S_{ij} combination and are depicted in Figure 6.6 for the Z resonance and the ZH training. Predictably, the strongest discrimination is between b-jets and s-, u-, d- jets and is roughly equivalent for all three background jets. The dominant background is from c-jets, originating from the similarity of b- and c-jets with a single reconstructed SV. Discriminating c-jets from u-, d- and s-jets exhibits similar performances, with relatively worse

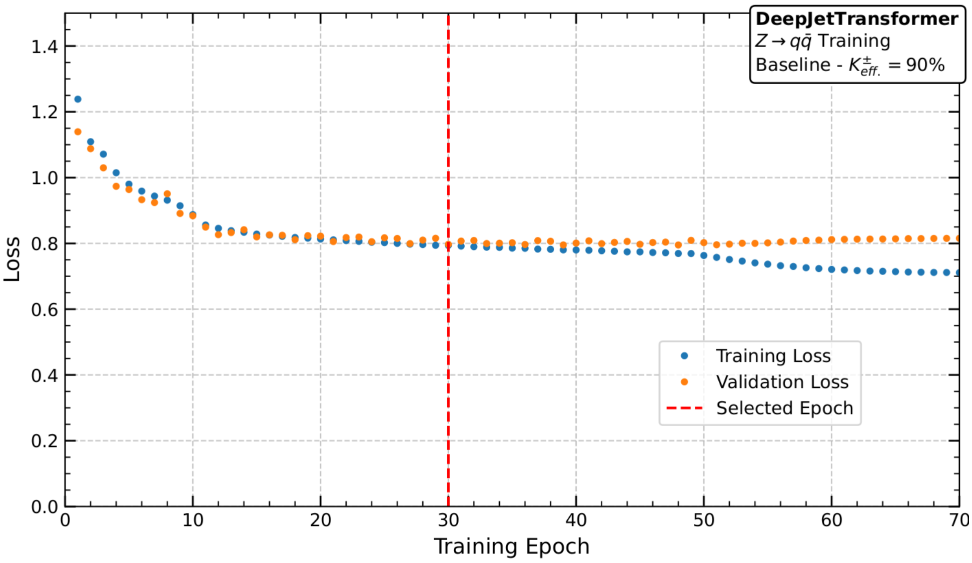


Figure 6.4: Training and validation loss curves as a function of epoch for the baseline DeepJetTranformer training on $Z \to q\bar{q}$ events at $\sqrt{s} = 91.2$ GeV.

discrimination of the s-jet background. Figure 6.6b shows that as the efficiency increases from the right to the left side of the plot, s-, u- and d-jets are discriminated worse than b-jets in the high-efficiency regime for c-jets until a turnover point at $\epsilon^{c}_{sig} \approx 80\%$, after which distinguishing s-, u- and d-jets becomes considerably easier than b-jets. Such a turnover can also be found in ParticleNetIDEA [59]. The sub-leading background comes from s-jets, clustered at low to mid charm scores, as also evident in Figure 6.5b, primarily as no SVs can be reconstructed for a significant number of c-jets, leaving few variables to distinguish c- and s-jets.

When s-jets are taken to be the signal, as shown in Figure 6.6c, c- and ud-jets present the most challenging backgrounds, with c-jets being easier to discriminate against at all signal purities. The c-jet background comes from jets where a charm hadron decays to a strange hadron, and only the V^0 can be reconstructed, or a strange hadron carries excess momentum. Some discrimination against the dominant ud-jets background can be achieved at higher cuts on the strange score, owing to the $K^{\pm}/\pi^{\pm}$ separation and V^0 reconstruction. Finally, Figures 6.5d and 6.5e show almost overlapping distributions of classifier scores for u- and d-jets. Figure 6.6d validates that classification is most challenging for u- and d-jets. When u-jets are taken to be the signal, it can be seen that DeepJetTransformer learns to discriminate u- vs d-jets with a $\epsilon^{u}_{sig} \approx 15\%$ at a $\epsilon_{bkg} = 10\%$, which is better than a random classifier, although not considerably. The discrimination is likely related to a mapping to the initiating parton's charge, such as the jet charge [164, 165], the effect of which is diluted by the presence of antiquarks.

While considering the performance for $H(\to q\bar{q})$ jets, depicted as dashed lines in Figure 6.6, no clear trend can be observed. Slight degradation in performance can be observed in the case of b tagging, compared to $Z \to q\bar{q}$ jets, particularly when c-jets are taken to be the background. The discrimination of c-jets vs s-, u-, and d-jets is found to perform relatively the best with respect to the $Z \to q\bar{q}$ jets when considering the percent-improvement in the ROC Area Under the Curve metric.

Figure 6.6f shows that the best quark-gluon discrimination can be achieved against the b quarks. This performance can be attributed to several discriminating variables, like jet-constituent multiplicity, constituent momentum distribution, etc., but is dominated by the presence or absence of reconstructed SVs. It is the most challenging to discriminate the s,

u, and d quarks from gluons due to their similar jet composition.

Qualitative Comparison with Other Taggers

A fair quantitative comparison with other taggers developed for future colliders is not feasible due to differing event samples and input features. However, the jet tagging performance trends are very similar to those of ParticleNetIDEA [59, 163]. The strange tagging efficiency of ParticleNetIDEA against the u-, d-jets surpasses that of DeepJetTransformer, owing to PID techniques like cluster counting and time-of-flight used by ParticleNetIDEA and the conservative PID estimates of DeepJetTransformer. A more detailed training dataset including such PID variables is expected to improve the tagging efficiencies of DeepJetTransformer.

DeepJetTransformer outperforms ParticleNetIDEA in bottom-gluon discrimination, especially for efficiencies lower than 90%. DeepJetTransformer also has a better discrimination of b-jet background for all other signal quark jet flavours. This efficient discrimination can be attributed to the inclusion of SVs.

6.5.1 Dependence on Particle Identification

Several $K^{\pm}$ classification scenarios were defined by fixing the efficiency of misidentification to $\pi^{\pm}$ and varying the $K^{\pm}$ identification efficiency. In addition, the limiting cases of Kaon identification with 0% and 100% efficiencies were considered. These are referred to henceforth as the no $K^{\pm}$ID and the perfect $K^{\pm}$ID scenarios. The considered efficiencies and the misidentification rates are the following:

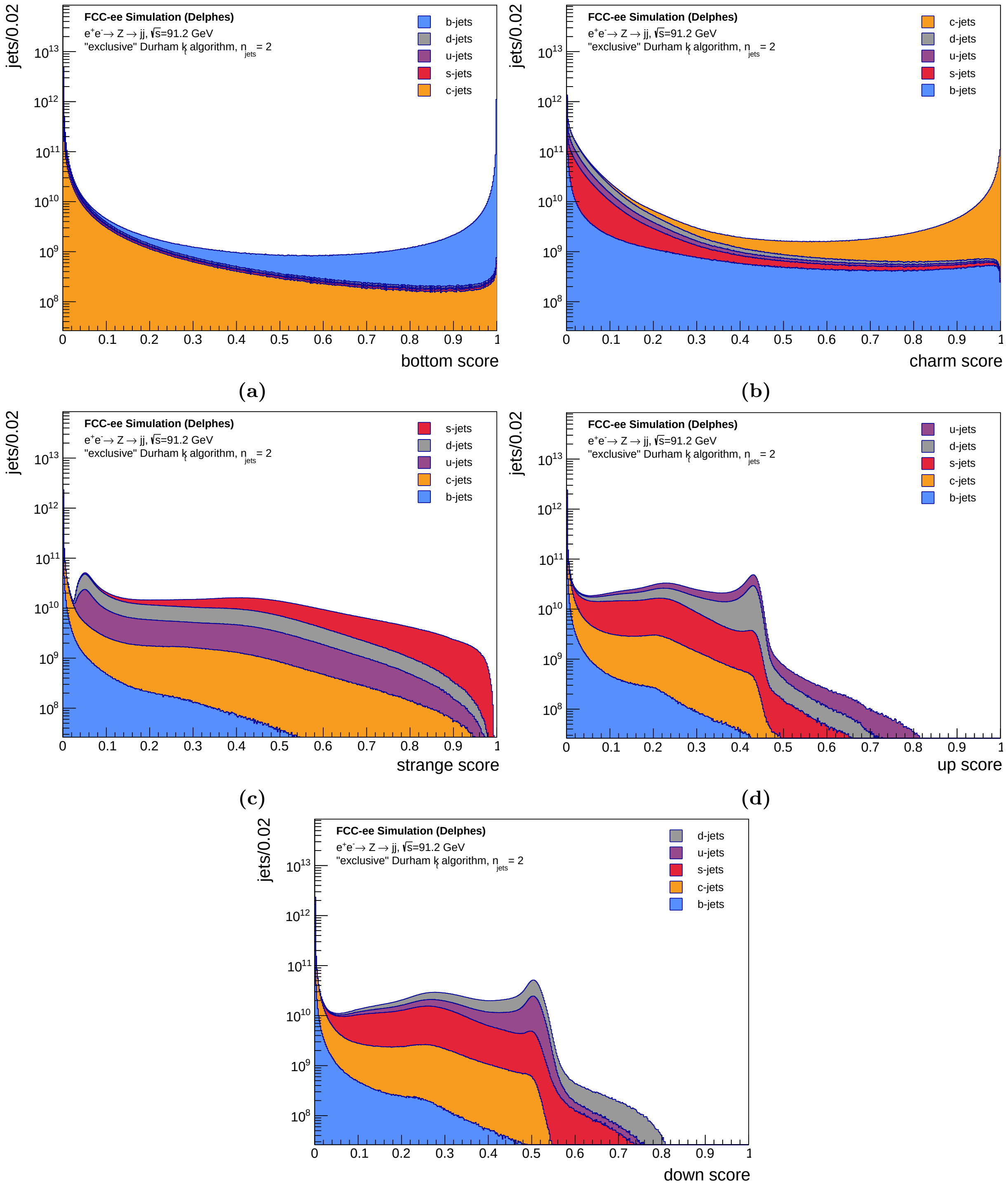
Table 6.4: Set of $K^{\pm}$ and $\pi^{\pm}$ particle identification points used to evaluate classification performance.

$K^{\pm}$ ID efficiency	0%	20%	40%	60%	80%	90%	95%	100%
$\pi^{\pm}$ misID efficiency	0%	10%	10%	10%	10%	10%	10%	0%

The no $K^{\pm}$ID scenario is used as the reference in this section to assess the impact of adding PID variables as input features for jet flavour tagging. The largest performance gain with the addition of $K^{\pm}$ID information is predictably in the classification of s vs ud jets, shown in Figure 6.7. Relative to the reference no $K^{\pm}$ID scenario, with a ϵ_{sig} of 31.6% at a ϵ_{bkg} of 10%, strange tagging efficiency improvements of 11.4%, 25.9%, and 32.9% are evident as the $K^{\pm}$ID efficiency is increased to 60%, 90%, and 95%, respectively. The perfect $K^{\pm}$ID scenario shows the most sizeable performance gain in ϵ_{sig} of 82.9%. This large performance improvement over the 95% $K^{\pm}$ID efficiency with the efficiency of misidentification to $\pi^{\pm}$ of 10% scenario suggests that minimising this misidentification is crucial to tagging strange jets, given their high $\pi^{\pm}$ multiplicity [156].

The performance gain for other forms of classification was marginal, with the exception of c vs ud and u vs d discrimination. For c vs ud, a performance gain of 1.8% from a ϵ_{sig} of 89.3% to 90.9% at a ϵ_{bkg} of 10% is observed while comparing the no $K^{\pm}$ID and the perfect $K^{\pm}$ID scenarios. In the case of u vs d, a 12.5% performance gain from a ϵ_{sig} of 13.6% to 15.3% at a ϵ_{bkg} of 10% is observed.

These results confirm the importance and necessity of particle identification techniques, especially for strange quark studies, as was also noted by some previous studies [59, 166, 167].



(a)

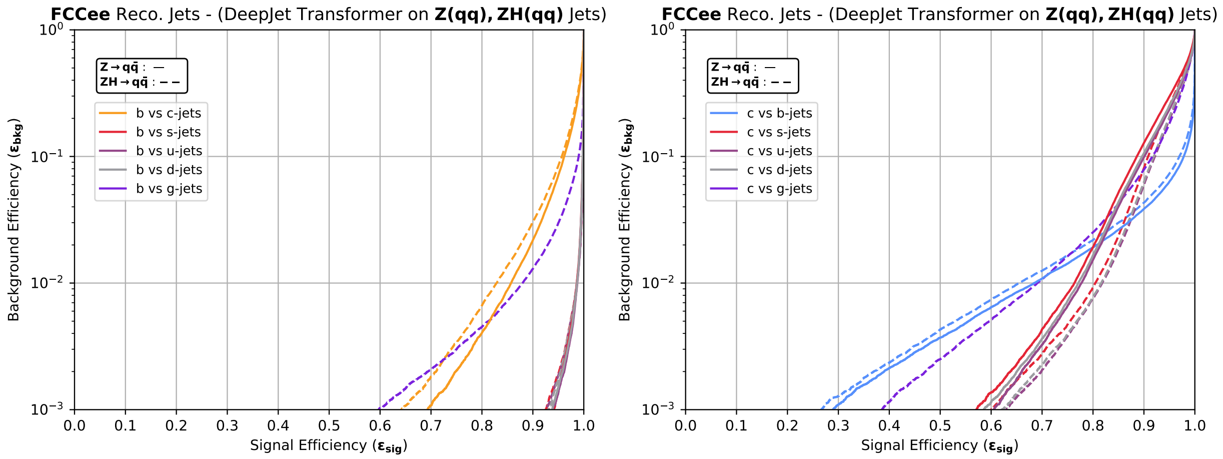
(b)

(c)

(d)

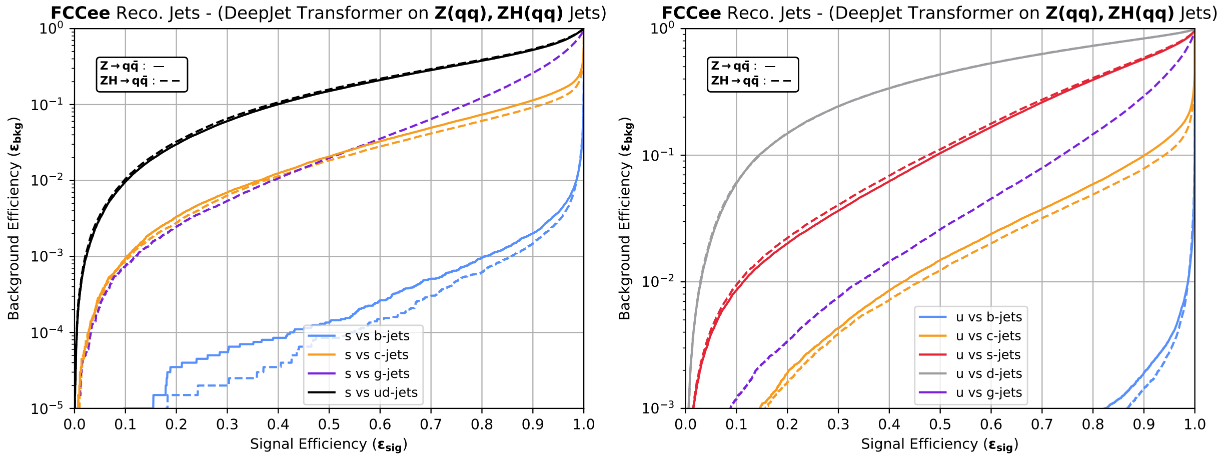
(e)

Figure 6.5: The softmaxed classifier outputs (S_i) of the bottom (a), charm (b), strange (c), up (d), and down (e) output nodes of DeepJetTransformer for $Z \rightarrow q\bar{q}$ events at $\sqrt{s} = 91.2$ GeV. The distributions are sub-divided by MC flavour.



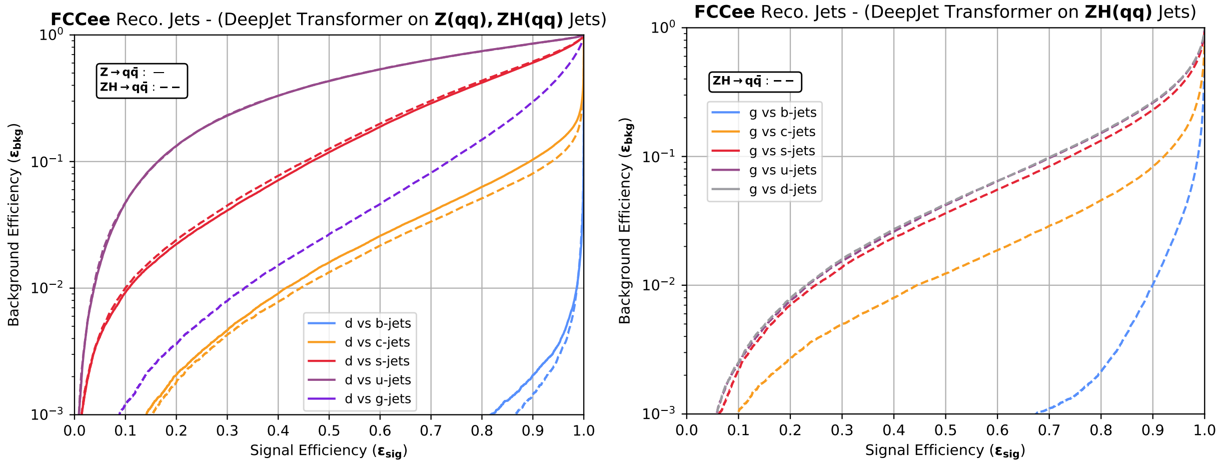
(a) bottom tagging

(b) charm tagging



(c) strange tagging

(d) up tagging



(e) down tagging

(f) gluon tagging

Figure 6.6: ROC curves for each S_{ij} combination, as defined in Eq. 6.4, where i is the signal parton flavour and j is the background flavour. The solid lines correspond to the classification of jets at the Z resonance at $\sqrt{s} = 91.2$ GeV, while the dashed lines correspond to the classification of jets from $Z(\to \nu\nu)H(\to q\bar{q})$ events at $\sqrt{s} = 240$ GeV. The tagger was trained separately for each process.

As noted earlier, an excess of V^{0}s, reconstructed K^0_S and Λ^0, carrying the bulk of the jet momenta is also a distinguishing feature of strange jets and these are expected to be more significant in the scarcity of charged Kaons. The inclusion of V^0 variables, as Figure 6.7 shows, results in an improvement of signal efficiency ranging from 14.3% in case of no $K^\pm$ID to 4.2% in the case of perfect $K^\pm$ID at a background efficiency of 10% for s vs ud discrimination. The percent improvement in signal efficiency for each of the $K^\pm$ID scenarios listed in Table 6.4 is depicted separately in Figure 6.7b. This trend proves the importance of V^{0}s to identify strange jets with low $K^\pm$ multiplicities or substandard $K^\pm/\pi^\pm$ discrimination. The performance gain in other forms of classification was again marginal.

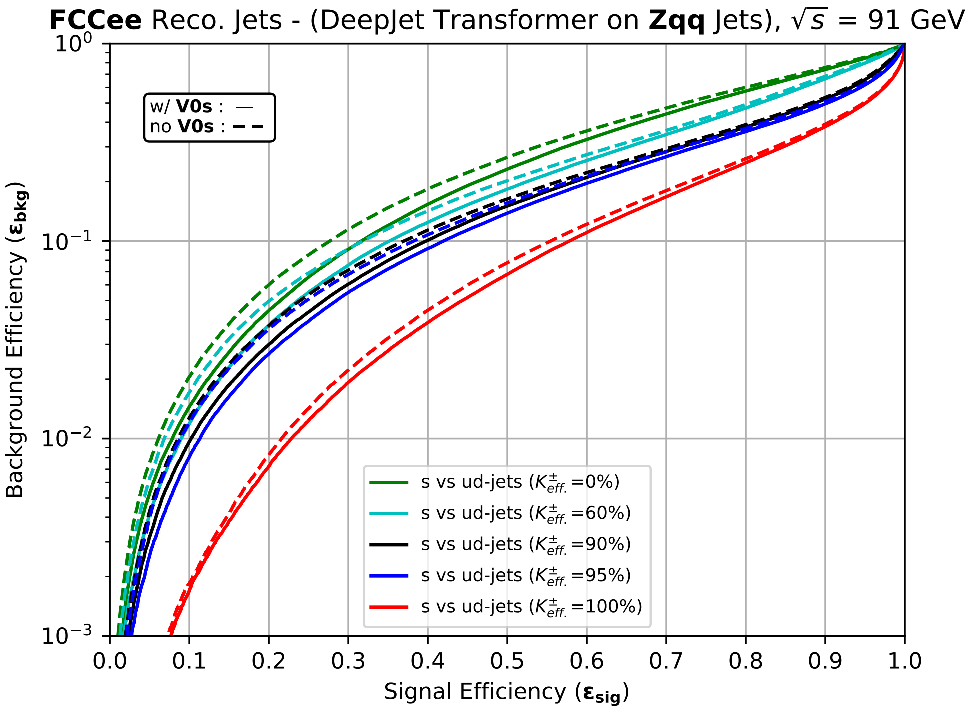
6.6 Feature Importance

Aiming to estimate the relative importance of a given variable class (e.g. SV variables) in the task of classifying jets, the classifier performance was evaluated using the Permutation Feature Importance [168, 169] method. For a jet i, let the variable classes introduced in Section 6.2 be denoted by $\{\vec{J}_i, \vec{C}_i, \vec{N}_i, \vec{S}_i, \vec{V}_i\}$. Then for the given variable class under investigation, for instance $\vec{S}$, the variable class was shuffled amongst all jets in the test data set, such that the jet i becomes

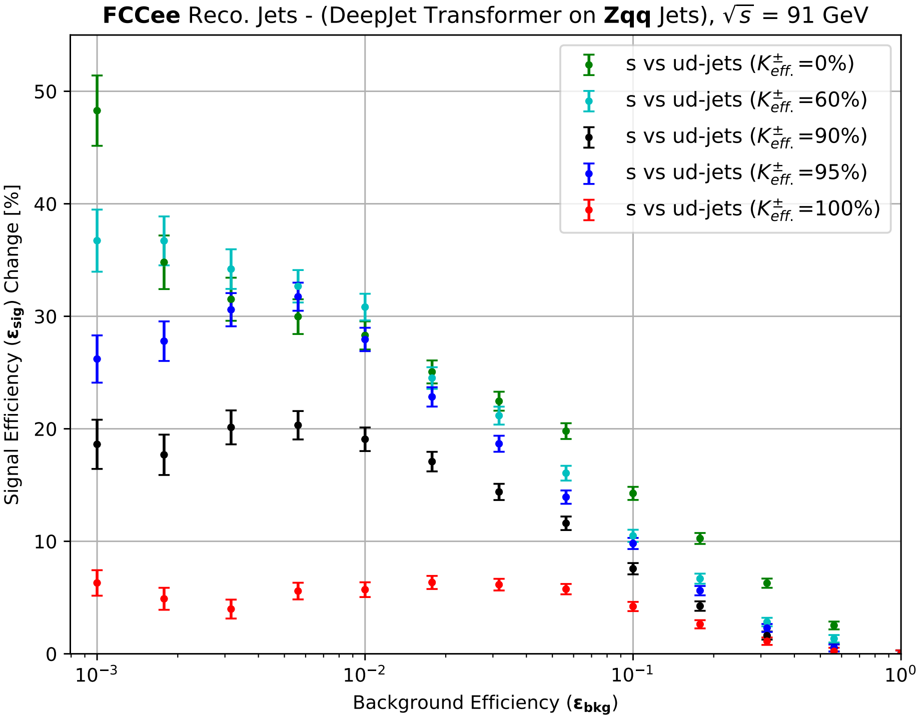
$$\{\vec{J}_i, \vec{C}_i, \vec{N}_i, \vec{S}_i, \vec{V}_i\} \rightarrow \{\vec{J}_i, \vec{C}_i, \vec{N}_i, \vec{S}_j, \vec{V}_i\} \tag{6.5}$$

where the variable class $\vec{S}$ now corresponds to that of some randomly selected jet j. In this way, the values for the variable class under investigation were randomly permuted across all jets in the dataset, disrupting the relationship between the permuted variable class and jet classification. This allows for an estimate of how much the performance of the classifier depends on the given variable.

The resulting performance change was considered for discriminating between b- vs c-, c- vs s-, and s- vs ud- jets, compared to the baseline where no variable classes were permuted, with respect to the change in signal efficiency. Charged jet constituent variables, listed in Table 6.2, were found to be the most impactful variable class for all types of discrimination at a background efficiency of $\epsilon_{bkg} = 10\%$, as detailed in Table 6.5. This is presumably due to more precise and discriminating information available for charged particles, arising from high-resolution tracking and vertexing measurements. SV variables, listed in Table 6.3, primarily benefited c vs s discrimination, with s vs ud tagging particularly insensitive. Of the remaining three variable classes, V^0 variables and neutral jet constituent variables were found to almost exclusively impact the performance of s vs ud discrimination, with little impact on both b vs c and c vs s discrimination, justifying the inclusion of V^{0}s for identifying s-jets through conservation of strangeness. Jet-level variables were found to be the least significant, marginally impacting s vs ud discrimination, and having virtually no impact on heavy flavour discrimination. While jet-level variables, including jet angularities, reflect differences in the initiating parton, it can be expected that sophisticated architectures like DeepJetTransformer capture these differences intrinsically by constructing mappings involving other variable classes. Moving to the high purity regime at a background efficiency of $\epsilon_{bkg} = 0.1\%$, primarily the same trends were observed, with the impact of any variable type being amplified. SV variables, in particular, became hugely important to heavy flavour tagging, reaching almost equal in impact to the charged jet constituent variables, proving that the presence and properties of SVs are definitive indicators for identifying heavy flavour jets.



(a) ROC curves



(b) Percent change in signal efficiency

Figure 6.7: The dependence of strange jet tagging performance on the inclusion of V^0s and charged Kaon identification scenarios. ROC curves for s vs ud tagging at the Z resonance at $\sqrt{s} = 91.2$ GeV (a). Solid lines represent results with the inclusion of V^0s, while dashed lines show the results without them. Percent change in signal efficiency (ϵ_{sig}) with the inclusion of V^0s for s vs ud tagging for each of the $K^{\pm}$ID scenarios listed in Table 6.4 (b). The axes are swapped with respect to Figure 6.7a to present the percent change in signal efficiency (ϵ_{sig}) as a function of 12 fixed background efficiencies (ϵ_{bkg}).

Table 6.5: Performance decrease in signal efficiency (ϵ_{sig}) after permutation of variable classes defined in Section 6.2 for fixed background efficiencies (ϵ_{bkg}) of 10% and 0.1%.

Variable Class		Jet-level	Charged	Neutral	SV	V^0
$\boldsymbol{\epsilon_{bkg} = 10\%}$	b vs c	2.4%	62.4%	2.2%	13.9%	0.1%
	c vs s	1.2%	65.7%	2.9%	29.6%	0.2%
	s vs ud	7.6%	59.4%	21.8%	5.0%	16.4%
$\boldsymbol{\epsilon_{bkg} = 0.1\%}$	b vs c	6.6%	97.0%	8.0%	89.9%	0.6%
	c vs s	9.3%	96.1%	11.0%	77.9%	0.2%
	s vs ud	35.9%	91.0%	57.3%	7.4%	43.8%

The above studies were repeated to estimate the relative importance of individual variables (e.g. m^{SV}), where rather than shuffling an entire variable class amongst jets, one individual variable was shuffled amongst itself. Adopting the same notation as for Equation 6.5, the jet-level variables of jet i can be written out, such that the variable classes become $\{\vec{J}_i, \vec{C}_i, \vec{N}_i, \vec{S}_i, \vec{V}_i\} = \{\vec{J}_i, \vec{C}_i, \vec{N}_i, \{s_1^i, ..., s_{25}^i\}, \vec{V}_i\}$. Then for the k^{th} SV variable (e.g. m^{SV}), the s_k is permuted amongst all other jets

$$\{\vec{J}_i, \vec{C}_i, \vec{N}_i, \{s_1^i, ..., s_k^i, ..., s_{25}^i\}, \vec{V}_i\} \rightarrow \{\vec{J}_i, \vec{C}_i, \vec{N}_i, \{s_1^i, ..., s_k^j, ..., s_{25}^i\}, \vec{V}_i\} \tag{6.6}$$

where the variable s_k^j now corresponds to that of some randomly selected jet j. The 64 variables included in DeepJetTransformer can be loosely split into the following categories:

- Kinematic ($|p|$, E, $|p|/|p|_{\mathrm{jet}}$, θ, $\Delta\theta$, …)
- PID ($isPhoton$, $K^{\pm}$ID, …)
- Track (D_0, z_0, …)

At a background efficiency of 10%, it was found that kinematic variables of charged particle constituents, including $\frac{E_{\mathrm{ch.}}}{E_{\mathrm{jet}}}$ and $\frac{|p|_{\mathrm{ch.}}}{|p|_{\mathrm{jet}}}$, were generally impactful, particularly for c vs s discrimination. Track variables, such as D_0/σ_{D_0} and z_0, were the most impactful, though less for b vs c than other types of discrimination, possibly due to their redundant information after the inclusion of SVs. PID variables had little impact on b vs c and c vs s discrimination, but $K^{\pm}$ID and photon ID were the most important for s vs ud discrimination, as was observed earlier. The high purity regime at a background efficiency of 0.1% resulted in similar trends, though with PID variables, including $K^{\pm}$ID and photon ID, decreasing in importance and being somewhat replaced by kinematic ones. It should be stated that the baseline $K^{\pm}$ID scenario, as mentioned in Section 6.2, is deliberately pessimistic, which could account for its decrease in importance. Track variables remained the most impactful. The secondary vertex mass m^{SV} became the most impactful variable in b vs c discrimination at high purity by a sizeable margin, as SV kinematics store essential information about the decaying hadrons. The results of this study are summarised in Table 6.6 below.

6.7 Z Peak Extraction

The decay of the Z boson was chosen for the evaluation of DeepJetTransformer due to the relative uniformity of its hadronic branching fractions, as well as the exquisite statistics that are expected from the FCC-ee's 4 year run at the Z-pole. Isolating $Z \rightarrow s\bar{s}$ events is a challenging, yet promising, avenue that has largely remained unexplored, in favour of leptonic or heavy-flavour channels, with their distinct signatures [158]. In this section the extraction of $Z \rightarrow s\bar{s}$

Table 6.6: Performance decrease in signal efficiency (ϵ_{sig}) after permutation of individual variables defined in Section 6.2 for fixed background efficiencies (ϵ_{bkg}) of 10% and 0.1%. A set of seven variables, chosen among the most impactful, is presented here.

Variable		$\ln(E_{\mathrm{ch.}})$	`isPhoton`	$K^{\pm}$ID	m^{SV}	$\lvert p\rvert^{\mathrm{V}^0}$	z_0	D_0/σ_{D_0}
$\epsilon_{bkg} = 10\%$	b vs c	3.5%	0.3%	0.2%	3.0%	0.1%	7.8%	11.6%
	c vs s	23.8%	0.7%	0.5%	0.3%	0.2%	20.9%	39.1%
	s vs ud	12.8%	16.6%	38.8%	0.0%	9.2%	23.3%	26.7%
$\epsilon_{bkg} = 0.1\%$	b vs c	13.8%	1.3%	0.9%	67.2%	0.8%	34.1%	45.0%
	c vs s	57.6%	0.9%	4.8%	7.0%	0.3%	56.2%	79.5%
	s vs ud	35.0%	28.0%	59.0%	0.4%	34.7%	60.5%	80.1%

events from $Z \to q\bar{q}$ events at the FCC-ee is reported, providing a realistic scenario for evaluating DeepJetTransformer's performance in discriminating against b, c, u, d-jets, simultaneously. Backgrounds beyond the hadronic jets from Z decays are not considered, but are expected to constitute less than 1% of events at the Z-pole. The events used in this study originate from the same $Z \to q\bar{q}$ at $\sqrt{s} = 91.2$ GeV sample introduced in Section 6.1. Events were required to consist of exactly two jets of matching MC flavour. Jets were required to pass fiducial cuts corresponding to a jet 3-momentum of at least $|p| > 20$ GeV and a polar angle in the range $14^\circ < \theta < 166^\circ$, ensuring the jet axis was sufficiently distant from the beam pipe. If either jet from an event did not satisfy the requirements, the event was rejected.

The adopted strategy for isolating s-jets consisted of first defining a s vs bc classifier according to Equation 6.4, where the individual scores for each class S_i were computed from the DeepJetTransformer softmaxed output nodes. It is important to note that the score for each $Z \to q\bar{q}$ jet was computed individually, with no knowledge of the event propagated to the network, as in previous sections in this chapter. The s vs bc classifier was evaluated at a set of background efficiencies $\epsilon^{b,c}_{bkg}$ of 10%, 1%, and 0.1%. Jets passing the s vs bc cut were subsequently subjected to a second cut by a s vs ud classifier defined similarly. The s vs ud classifier was likewise evaluated at a set of background efficiencies $\epsilon^{u,d}_{bkg}$ of 10%, 1%, and 0.1%, now defined with respect to the set of jets passing the first cut. Four working points (WP) of increasing purity were defined by considering different combinations of background efficiencies, or mistag rates, for the sequential classifiers. The working points, along with their respective efficiencies, are summarized in Table 6.7, where it can be seen that while the first classifier removes the comparatively "easy" background of heavy flavour jets while maintaining a high signal efficiency, even at the 1% working point. The s vs ud classifier, however, shows a sharp decrease in signal efficiency when moving to more stringent working points, as discrimination against ud-jets relies heavily of Kaon identification, and the baseline version of the DeepJetTransformer makes relatively modest PID assumptions. Interestingly, the signal efficiency of the s vs ud classifier only varies from $\epsilon^{s}_{sig} = 40.03 \pm 0.04\%$ to $\epsilon^{s}_{sig} = 39.28 \pm 0.06\%$, demonstrating the apparent orthogonality of the classification problem.

The Z resonance depicted on Figure 6.8a was obtained as the invariant mass of the sum of the 4-momenta of both jets in an event $p^\mu = (p_1)^\mu + (p_2)^\mu$: $m = \sqrt{p^\mu p_\mu}$. Events are split by flavour, with light jets appearing roughly symmetrically about ~ 90 GeV. Z boson decays to heavy quarks display a pronounced tail towards lower mass corresponding to energy carried away by neutrinos arising from semi-leptonic decays of the hadrons containing b and c quarks. Figure 6.8b depicts the dijet invariant mass for all events passing the s vs bc cut for WP2/3. Figure 6.9a depicts the dijet invariant mass for all events which additionally pass the s vs ud cut

Table 6.7: Signal efficiency at the four working points defined with respect to mistag rates of the s vs bc and s vs ud cuts. The number of signal events N_{sig} and background events N_{bkg} is listed explicitly for an integrated luminosity of 125 ab^{-1}. Signal efficiencies and mistag rates are defined with respect to $Z \to q\bar{q}$ jets.

		Mistag Rate [%]	Efficiency [%]	N_{sig}	N_{bkg}
WP1	s vs bc	10	98.93 ± 0.03	7.35×10^{11}	1.35×10^{12}
	s vs ud	10	40.03 ± 0.04	1.45×10^{11}	3.25×10^{10}
WP2	s vs bc	1	54.18 ± 0.04	2.38×10^{11}	2.06×10^{11}
	s vs ud	10	39.28 ± 0.06	5.10×10^{10}	5.57×10^{9}
WP3	s vs bc	1	54.18 ± 0.04	2.38×10^{11}	2.06×10^{11}
	s vs ud	1	10.05 ± 0.11	1.12×10^{10}	4.77×10^{8}
WP4	s vs bc	0.1	17.96 ± 0.06	3.23×10^{10}	6.98×10^{9}
	s vs ud	0.1	1.98 ± 0.33	3.56×10^{8}	3.38×10^{6}

defined sequentially for WP3 in Table 6.7. The Z peak is extremely pure, with the remaining background consisting almost exclusively of u- and d-jets. The precise number of signal and background events passing the WP3 cuts is listed on Table 6.7 for an integrated luminosity of 125 ab^{-1}. While this study does not consider machine backgrounds or irreducible backgrounds from other Standard Model processes, both expected to be at the per cent level, it is nevertheless instructive to consider the Asimov estimate of the discovery significance [170, 171], defined as

$$Z = \sqrt{2\left[(N_{sig} + N_{bkg}) \log\left(1 + \frac{N_{sig}}{N_{bkg}}\right) - N_{sig}\right]}. \tag{6.7}$$

Figure 6.9b depicts the discovery significance as a function of integrated luminosity for the different working points. A discovery significance corresponding to the canonical 5 σ is reached extremely quickly for all four of the working points. WP3, in particular, reaches 5 σ after an integrated luminosity of only 60 nb^{-1}, corresponding to less than a second of the Z-pole run at the FCC-ee.

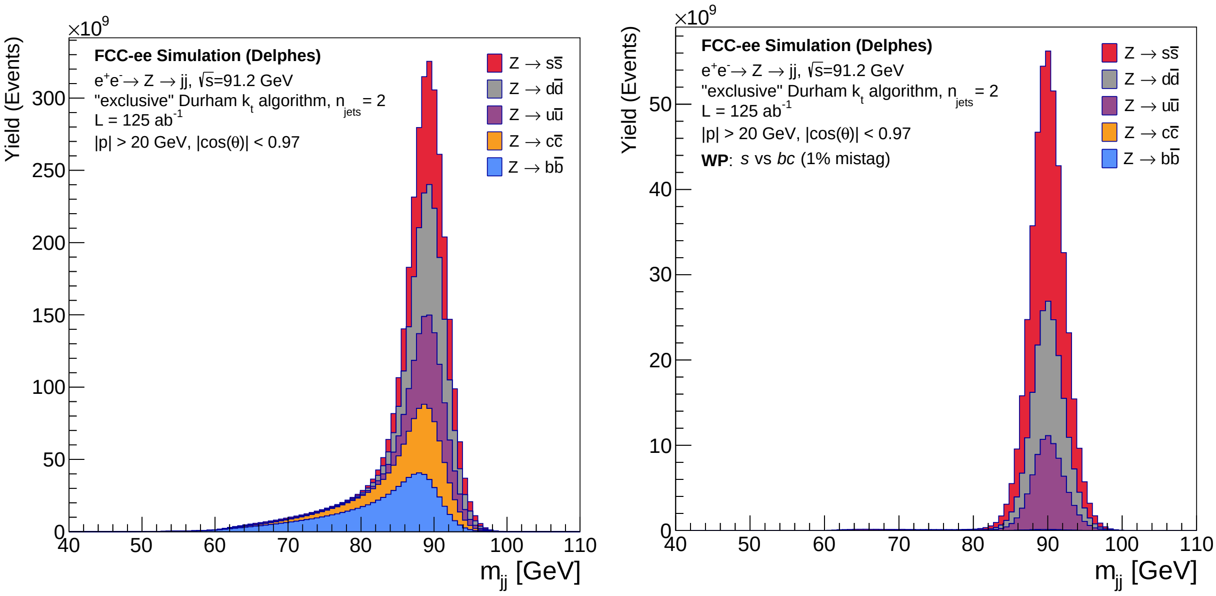


Figure 6.8: Dijet invariant mass of $Z \to q\bar{q}$ events at $\sqrt{s} = 91.2$ GeV before any cuts (a), and for events where both jets pass the s vs bc cut of WP2/3 of Table 6.7. An integrated luminosity of 125 ab^{-1} is assumed.

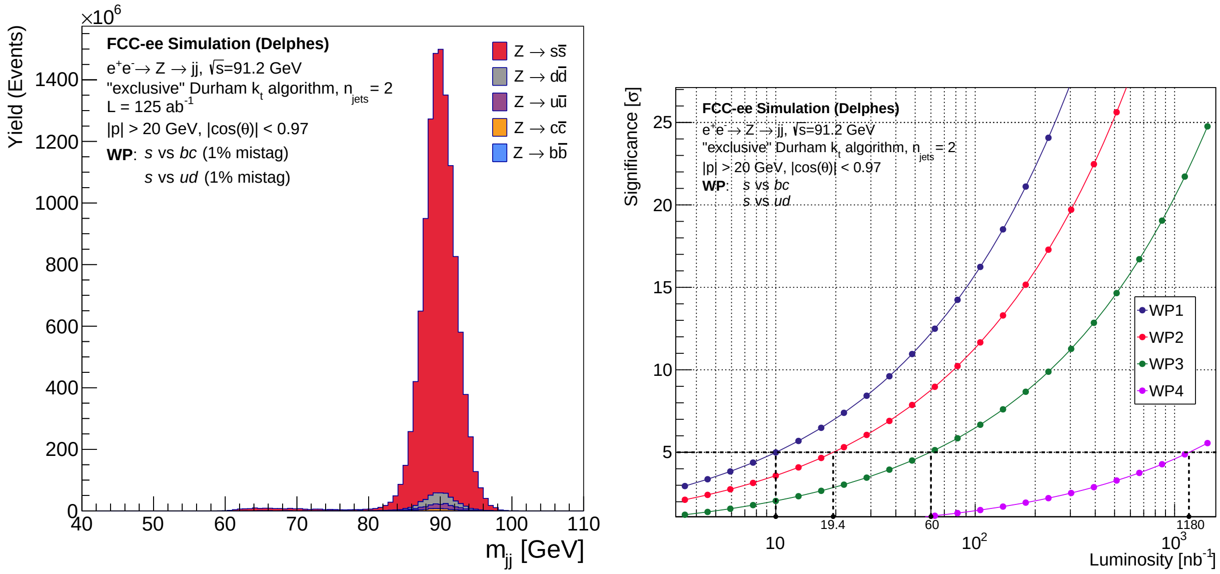


Figure 6.9: Dijet invariant mass of $Z \to q\bar{q}$ events at $\sqrt{s} = 91.2$ GeV where both jets pass the s vs bc and subsequent s vs ud cut of WP3 of Table 6.7 (a). An integrated luminosity of 125 ab^{-1} is assumed. Discovery significance of $Z \to s\bar{s}$ events at 91.2 GeV as a function of integrated luminosity for WP1-4. The dashed lines denote the luminosity required for a 5 σ significance for each of the WPs.

6.8 Conclusions

The discrimination of strange quark jets at the Z-pole run of the FCC-ee was demonstrated using a computationally-efficient, transformer-based neural network. This approach combines the representational power of the attention mechanism with a novel implementation of vertexing at the FCC-ee, achieving state-of-the-art performance. The impact of particle identification on strange tagging was explicitly explored through the inclusion of V^0 reconstruction and varying degrees of $K^\pm$ discrimination.

An excellent discrimination of b-, c-, s-jets was achieved, with signal efficiencies in the $\epsilon_{sig} \sim$ 90%+ range for both b vs c and c vs s tagging, at a background efficiency of $\epsilon_{bkg} \sim 10\%$. The model performance showed little dependence on the events, with jets originating from $Z \to q\bar{q}$ and $H \to q\bar{q}$ decays being discriminated with similar efficacy, though a slight improvement was observed in the Higgs sample for c-jet tagging specifically.

The u, d background proved to be the most challenging throughout these studies, particularly for s-jets, though this was mitigated by Kaon identification. Among the considered $K^\pm$ scenarios, the improvement was found to depend more on the $\pi^\pm$ misidentification than on increasing $K^\pm$ identification efficiency, reflecting the high pion multiplicity in light jets. Similarly, the reconstruction of V^0 particles was found to benefit strange jet tagging sizeably, though the improvement was reduced as the $K^\pm$ identification improved.

Subsequent permutation feature importance studies on b vs c, c vs s, and s vs ud tagging performance further emphasised the importance of Kaon identification for strange jet discrimination. For heavy flavour tagging, and even for s vs d at high purity, track and SV variables proved most impactful.

Finally, a case study targetting the isolation of $Z \to s\bar{s}$ jets using DeepJetTransformer at the Z-pole was performed, illustrating its use in a concrete physics scenario. It was shown that a discovery significance of 5 σ is reached after less than a second of the Z-pole run at the FCC-ee, albeit in a simplified scenario where non-Z backgrounds are neglected.

DeepJetTransformer offers a lightweight architecture for efficient discrimination of jets initiated by all quark flavours. The computational complexity, measured in FLOPs, is approximately 19.7 MFLOPs. Comparatively, DeepJetTransformer requires fewer FLOPs than competing architectures [94, 110], making it an excellent choice to efficiently evaluate the impact of the constantly evolving detector design on flavour tagging.

Future work may focus both on the optimisation of the machine learning algorithm, and the realism of input features. The network performance could benefit from refining the input feature set, hyperparameter tuning, and a larger training set. In parallel, more realistic $K^{\pm}$ PID scenarios using the time-of-flight or number of primary ionisation clusters, as in Ref. [59], could be incorporated. A detailed outlook for further developments is presented in Chapter 11 at the end of this thesis.

Ultimately, while increasingly sophisticated architectures may be able to extract additional discriminative power, the jet tagging performance will be gated by the quality of information that is available. These studies emphasise the central importance of particle identification to strange jet tagging at future colliders. The potential for further gains through precise vertexing enabling not only increasing precision in secondary vertex reconstruction, but also V^0 decays, motivates ongoing MAPS development and optimisation, as explored in the following chapters of this thesis.

Part III

Monolithic Active Pixel Sensors for Future Colliders

7 Silicon Detectors

Silicon detectors have emerged as the enabling technology in high-precision tracking applications since their adoption in the 1980s. Their introduction vastly improved the spatial reconstruction of particle tracks, leading to a more precise determination of particle properties such as momentum, and enabling the reconstruction of SVs for the first time. Secondary vertex reconstruction, and tracking in general, are critical components of detector-level information for a wide variety of measurements, and in particular the identification of jets. While novel jet tagging algorithms, such as the one introduced in Chapter 6, will continue to incrementally improve the achievable jet flavour tagging accuracy, the fundamental constraint will always remain the information on the particle trajectories and properties from which the algorithm must infer the initiating parton.

Owing to the low energy ($\sim$ 3.6 eV) required to create an electron-hole pair, silicon detectors can be regarded as extremely sensitive ionisation chambers, where energy deposited by an impinging particle is translated to an electrical signal by the generation and subsequent motion of charge carriers. The resulting signal, however, is a small current and must thus undergo a sophisticated signal processing chain in order to become practically usable, balancing a variety of design requirements including the spatial and energy resolutions, power density, material budget, radiation tolerance, cost, and many others. Naturally, a variety of different technologies, each optimal for a different environment or measurement, have emerged.

This chapter provides an introduction to silicon detectors, with a strong bias towards topics relevant to the characterisation of the monolithic active pixel test structures presented in this thesis. Much more comprehensive treatments of the topics covered can be found in Refs. [53, 172, 173]. The chapter begins with the fundamental properties of semiconductors, followed by an overview of *pn*-junctions. Subsequently, Section 7.2.1 explores the energy deposition of charged particles and photons in matter. The signal formation and its readout are presented in Sections 7.2.2 and 7.3.1, respectively. Noise sources and their impact on signal detection are presented in Section 7.3.2. Finally, Section 7.4 provides a brief overview of monolithic active pixel sensors and their applications.

7.1 Semiconductors

Solid materials can be classified according to their conductivity into insulators, conductors, and semiconductors. The close spacing of atoms leads to overlapping of the individual wavefunctions of atoms, leading to a splitting of the energy levels [53]. For a large number of atoms typical of a solid, this results in an almost continuous *band* of allowed energies. The electrical conductivity of a solid is determined by two of these bands: the valence band and the conduction band. The *valence* band is the highest energy band that is completely filled at 0 K. The valence band is separated from the *conduction* band by a finite energy gap covering a range of "forbidden" energies, as depicted on Figure 7.1. The size of the band gap defines whether a material is considered an insulator or a semiconductor. While this distinction is not strict, typical band gap values include $\sim 0.5 - 3$ eV for semiconductors and > 4 eV for insulators.

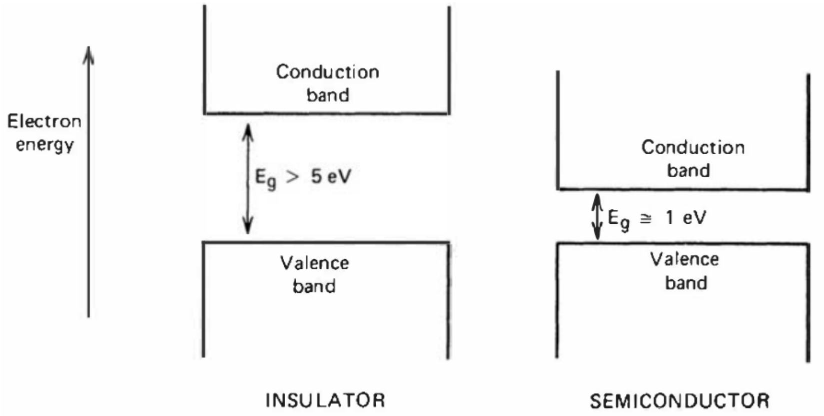


Figure 7.1: Schematic of the bandgap structure of an insulator and a semiconductor. The bandgap energy is denoted by E_g. Figure taken from Ref. [173].

The *Fermi energy* E_F, or chemical potential, describes the highest allowable energy level for an electron at 0 K. In the case of insulators and semiconductors the Fermi energy lies near the middle of the band gap. Thus, the valence band is completely filled, while the conduction band is completely empty for both insulators and semiconductors at 0 K, preventing the conduction of any current. Above 0 K, the behaviour of insulators and semiconductors begins to differ. Electrons can be thermally excited into the conduction band of semiconductors. The probability for an electron to occupy a higher energy state is given by [172]

$$f_e(E) = \frac{1}{e^{(E-E_F)/k_BT} + 1} \tag{7.1}$$

where k_B is the Boltzmann constant. While the probability of an individual electron occupying an energy state in the conduction band is negligibly small, the high density of states leads to a non-trivial conductivity for semiconductors. While this in principle also occurs in insulators, the band gap of insulators is large enough that the concentration of electrons in the conduction band is negligible.

In the case of silicon, the thermal excitation of electrons corresponds to overcoming the energy of the valence bond with the neighbouring atom, and "freeing" the electron. This process generates an electron which now acts a free charge carrier, and a hole in the valence bond in which the electron once partook. The hole, as the electron, contributes to overall conductivity of silicon, and through the collective motion of neighbouring electrons behaves as a free charge carrier, albeit with a reduced mobility with respect to the electron. The concept of *mobility* μ describes how quickly a charge carrier moves in a material under the influence of an electric field. At low electric fields the relationship is roughly linear [173]:

$$v_{e,h} = \mu_{e,h} \cdot |\vec{E}| \tag{7.2}$$

where $v_{e,h}$ denotes the drift velocity of electrons and holes, respectively.

The conductivity of a semiconductor is given as

$$\sigma = e(n \cdot \mu_e + p \cdot \mu_h) \tag{7.3}$$

where n, p and μ_n, μ_h denote the electron and hole densities and mobilities, respectively. The electron and hole densities are related by the *law of mass action*, which states that the product of the charge carrier densities in a semiconductor is constant at equilibrium:

$$n \cdot p = n_i^2 \tag{7.4}$$

where n_i is intrinsic carrier density. In an *intrinsic* semiconductor like pure silicon, the thermal generation of charge carriers, and their potential recombination, will by construction generate an equivalent number of electrons in the conduction band and holes. Thus, the free electron and hole concentration will be equal, simplifying Equation 7.3 to $\sigma = en_i(\mu_e + \mu_h)$.

Doping

Extrinsic semiconductors denote semiconductors to which an impurity has been introduced. The introduction of impurity atoms into the pure semiconductor lattice is known as *doping*, and entails an increase in the conductivity of the respective semiconductor. In n-type silicon pentavalent atoms such as phosphorus are introduced to the silicon lattice. Four of the valence electrons form covalent bonds with neighbouring silicon atoms, while the remaining electron remains in a quasi-free state. The remaining electron can be easily excited into the conduction band by thermal excitations as it is not tightly bound in a covalent bond. It occupies an energy level slightly below the conduction band in the band gap, as depicted on Figure 7.2a. While the law of mass action still holds, the concentration of electrons in the conduction band begins to be dominated by electrons from the impurity atoms. In addition, the sharp increase in electrons in the conduction band leads to a recombination with remaining holes, such that $n >> p$. Equation 7.3 reduces to $\sigma = e \cdot n \cdot \mu_e$, as the conductivity of the material begins to be dominated by the majority charge carrier, in this case electrons.

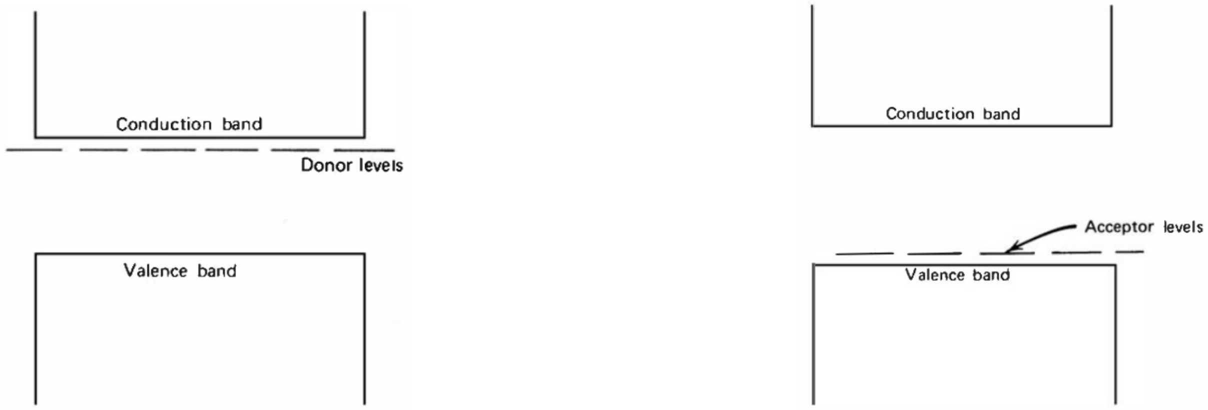


Figure 7.2: Schematic of the modified bandgap structure of an n-doped (a) and p-doped (b) semiconductor. Figures taken from Ref. [173].

Similar doping can be performed with trivalent atoms, such as boron, which instead of introducing an excess electron into the lattice, have one too few. The doping, known as p-type doping, entails a new energy level which appears slightly above the valence band, as depicted on Figure 7.2b. Electrons in the conduction band preferentially recombine with the holes from the impurity atoms, as the electrons are less tightly bound than in the Si-Si valence bonds. Analogously to n-type doping, the holes now become the majority charge carriers, and the conductivity likewise increases. Typical doping concentrations in Silicon are in the range 10^{12}-10^{18} cm^{-3} [172]. In semiconductor detectors doping is used either for implants or for bulk material. *Implants* refer to small regions of highly doped silicon which can act as ohmic contacts (e.g.

for the collection electrode). The high doping ensures a low resistivity, and such regions are accordingly denoted as p^+, n^+ depending on the type of impurity. By contrast, the *active region* where electron-hole pairs are generated is often doped with at much lower concentrations, as the higher resistivity is beneficial to charge collection, as detailed in the following section. The low doping is denoted by p^-, n^-.

pn-junctions

Combining two oppositely-doped regions of silicon, typically by introducing n- and p-type impurities to different volumes of the same silicon crystal, results in the creation of a *depletion region*, which has several desirable properties for particle detection. Figure 7.3 depicts the interface between the n- and p-type regions, referred to as a pn-junction. The formation of the depletion region in a pn-junction begins with the majority carriers in the n-type silicon diffusing into the p-type silicon, and vice-versa. This leads to an electron-hole pair recombination, whereby conduction electrons from the n-type silicon form valence bonds with holes in the p-type silicon. In this way both atoms in the n-type and p-type silicon are ionised, and the depletion region, which is devoid of charge carriers, begins to form. While the pn-junction as a whole is electrically neutral, the n-type impurity atoms are positively charged, while the p-type impurity atoms are negatively charged. Accordingly, a potential difference across the depletion region resulting from the oppositely charged ions also forms. The resulting electric field opposes the flow of charge carriers along the diffusion gradient, with conduction electrons being swept back to the now positively-charged n-type silicon, and vice-versa. At equilibrium, the motion of the charge carriers due to diffusion and drift along the electric field are equal and opposite, such that the depth of the depletion region remains constant. Although electron-hole pairs continue to be thermally-generated within the depletion region, the electric field causes them to drift to their corresponding n- or p-type silicon, before any meaningful build-up of charge carriers.

The depletion region is central to the functioning of semiconductor detectors, since the absence of thermally-generated charge carriers translates to greatly reduced noise. Moreover, charge carriers resulting from the energy deposition of particles traversing the detector material drift towards either the n- or p-type silicon generating a signal, as detailed in Section 7.2.2. By applying a reverse bias voltage, that is to say a voltage along the built-in potential V_{bi}, the depth of the depletion region can be increased, as depicted in the bottom panel of Figure 7.3. In particular, the electric potential $\varphi(\vec{r})$ satisfies Poisson's equation:

$$\nabla^2 \varphi(\vec{r}) = -\frac{\rho(\vec{r})}{\epsilon} \tag{7.5}$$

where ϵ is the dielectric constant and $\rho(\vec{r})$ is the charge density. Assuming the potential depends only the direction perpendicular to the pn-junction, and noting that the charge density is given by the respective carrier concentrations for each of the junction sides, reduces Equation 7.5 to

$$\frac{d^2}{dx^2}\varphi(x) = \begin{cases} -\frac{eN_D}{\epsilon} & \text{if } 0 < x < x_n \\ \frac{eN_A}{\epsilon} & \text{if } -x_p < x < 0 \end{cases} \tag{7.6}$$

where $-x_p$ and x_n denote the boundaries of the depletion region, and $x = 0$ the junction interface. Given the constant charge density, Equation 7.6 can easily be integrated in order to obtain the x-component of the electric field in each respective region:

$$\frac{d}{dx}\varphi^{n,p}(x) = -E_x^{n,p}(x) = -\frac{\rho^{n,p}}{\epsilon} \cdot x + C_{n,p} \tag{7.7}$$

where the integration constant is determined from the boundary condition that the electric field vanishes at the boundaries, yielding

$$\frac{d}{dx}\varphi(x) = \begin{cases} -\frac{eN_D}{\epsilon}(x - x_n) & \text{if } 0 < x < x_n \\ \frac{eN_A}{\epsilon}(x + x_p) & \text{if } -x_p < x < 0 \end{cases} \tag{7.8}$$

The above Equation can be integrated again to obtain

$$\varphi(x) = \begin{cases} -\frac{eN_D}{2\epsilon}(x - x_n)^2 + V_{bias} & \text{if } 0 < x < x_n \\ \frac{eN_A}{2\epsilon}(x + x_p)^2 & \text{if } -x_p < x < 0 \end{cases} \tag{7.9}$$

where the integration constant is now determined from the boundary conditions of the applied voltage: $\varphi(x_n) = V_{bias}$ and $\varphi(-x_p) = 0$.

The continuity of the potential at the junction interface $\varphi(0^+) = \varphi(0^-)$ allows the expression of the bias voltage as a function of the depletion region depth of each region (x_n, x_p) using Equation 7.9. Using the charge neutrality across the junction $N_D \cdot x_n = N_A \cdot x_p$, the relation can easily be solved for the depth of the depletion region $(d = x_n + x_p)$, making its approximately square-root dependence on the bias voltage manifest [173]:

$$d \cdot x_p = \frac{2\epsilon}{N_A e} V_{bias} \tag{7.10}$$

which can be written as

$$d \approx \sqrt{\frac{2\epsilon}{N_A e} V_{bias}} \propto \sqrt{V_{bias}} \tag{7.11}$$

by letting $d \approx x_p$. This holds when the doping concentration of the n-type region is much higher than for the p-type region, as is the case for the junction between the n-well collection electrode and the lightly-doped p-type sensitive layer of the chips explored in this thesis. Notably, the depth of the depletion region also depends on the inverse root of the doping concentration of the lightly-doped layer, suggesting that the extent of the depletion region can be increased by minimising the doping concentration.

The extent of the depletion region is one of the defining characteristics in the performance of a semiconductor detector. In addition to defining the region in which generated charges will drift to the collection electrode, the width of the depletion region determines the junction capacitance of the detector. Treating the *pn*-junction as a parallel plate capacitor, a capacitance of

$$C = \frac{\epsilon A}{d} \approx A \cdot \sqrt{\frac{N_A \cdot e \cdot \epsilon}{2 \cdot V_{bias}}} \tag{7.12}$$

is obtained, where A is the area of the *pn*-junction interface. The capacitance should be minimised in order to induce a larger signal for an equivalent number of deposited charge. A reduction in capacitance can greatly mitigate electronic noise contributions arising from later stages of the signal processing, as described in Section 7.3.2.

The thermal generation of charge carriers can contribute to a current in the absence of a charge deposition from an ionising particle. The *leakage current* is increased substantially by impurities and radiation damage to the silicon bulk, which can introduce intermediate energy levels that

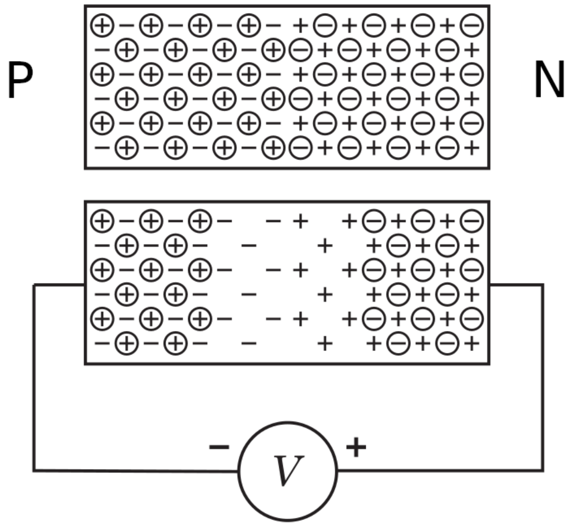


Figure 7.3: Schematic of a *pn*-junction consisting of a *p*-doped and an *n*-doped region. Free charge carriers are denoted by + and − signs, while the resulting ions are circled (top). The width of the depletion region is increased if a reverse bias voltage is applied, denoted by V. Figure adapted from Ref. [172].

facilitate excitations. The leakage current is a significant source of noise in semiconductor detectors, that is not mitigated by reducing the junction capacitance.

Although the discussion has focused exclusively on *pn*-junctions thus far, it is worth noting that p^+p- and n^+n-junctions, formed from the same type of impurity at different doping levels, share some similar properties with *pn*-junctions [53]. Assuming a p^+p-junction, as the interface between the active layer and the substrate for the chips studied in this thesis, the hole concentration will be much higher in the p^+ region leading to the diffusion of holes into the more lightly-doped p region. This causes the build up of negative ions in the p^+ region, and thus the build-up of an electric field opposing further diffusion, in analogy with a *pn*-junction. This field acts as a reflective barrier for electrons approaching the p^+ region due to the negative ion build-up [174].

7.2 Signal Formation

Fundamentally, the signal in a semiconductor detector consists of the electron-hole pairs created by impinging radiation. For instance, a 5.9 keV K_α photon from an ^{55}Fe decay will create a photoelectron as it impinges on silicon, which will in turn generate $\sim$ 1640 electron-hole pairs as it deposits its energy in the detector. More generally, the energy required to generate a single electron-hole pair in silicon is 3.6 eV, which is considerably larger than the band gap of 1.12 eV, as large fraction of the energy goes into phonon excitations [172]. Electrons and holes then drift along the electric field into the p and n-type region respectively, generating a signal, as described in Section 7.2.2. Before detailing the collection of said signal, some of the fundamentals of the energy deposition of particles as they traverse detector material will be presented.

7.2.1 Particle Interactions with Matter

The energy deposition of charged and neutral particles differs substantially due to distinct interaction mechanisms, and must thus be treated separately. In what follows, the interactions of electrons and heavier charged particles is discussed, followed by a brief overview of the photon interactions with matter, in line with measurements presented in Chapters 9 and 10.

Charged Particles

Charged particles traversing detector material deposit a fraction of their energy by ionising the atoms in the detector material. The mean energy deposited by charged particles in the range $0.1 \lesssim \beta\gamma \lesssim 1000$ per unit length is given by the Bethe-Bloch equation [158, 175]:

$$-\left\langle \frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \left(\frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} \right) - \beta^2 - \frac{\delta}{2} \right], \tag{7.13}$$

where $K = 4\pi N_A r_e^2 m_e c^2$, z is the charge of the impinging particle, Z, A are the atomic number and mass of the absorber, I is the mean excitation energy, W_{max} is the maximum possible energy transfer to an electron, and δ is a high energy correction related to the polarisation of the absorber [158]. The energy loss in a number of media is depicted on Figure 7.4a. The distribution is characterised by three regions. In the low-momentum range, primarily for $\beta\gamma < 1$, Equation 7.13 is dominated by the $1/\beta^2$ term. The high energy-loss can be associated with large momentum transfers that occur due to the long interaction times [53]. At even lower energies Equation 7.13 does not hold.

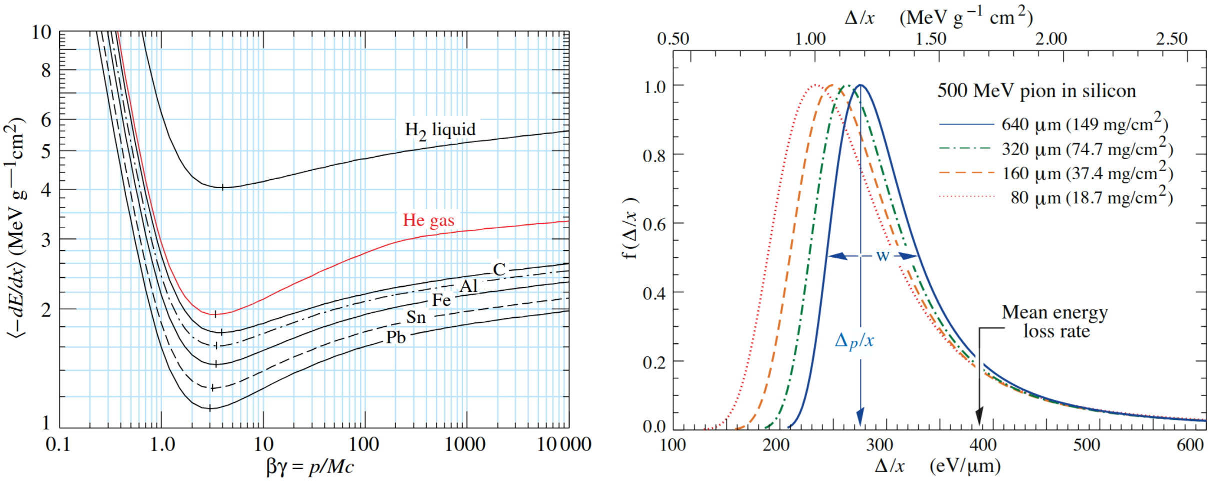


Figure 7.4: Mean energy loss in a range of media as a function of $\beta\gamma$ (a). Straggling functions of silicon for 500 MeV pions, shown for a range of absorber thicknesses, as a function of the energy loss per unit length Δ/x (b). Figures adapted from Ref. [158].

Around $\beta\gamma \approx 3$ the energy-loss reaches its minimum as the scaling from the $1/\beta^2$ term becomes diminished, but the high-momentum scaling of $\ln(\beta^2\gamma^2)$ has yet to kick-in. Particles in the $\beta\gamma \approx 3$ range are accordingly denoted *minimum ionising particles* (MIPs), though in practice, and throughout this thesis, the term is also used to include particles with higher momenta. While the Bethe-Bloch equation given in Equation 7.13 predicts a rise at high $\beta\gamma$ due to the scaling of the $\ln(\beta^2\gamma^2)$ term, before density corrections δ begin to flatten it, a fraction of the ionisation energy is often not measured as it is carried away from the detector by energetic electrons. This effect can be substantial, particularly for thin silicon detectors. The so called *delta electrons* arise from rare collisions where the energy deposition by the impinging charged particle is substantially higher than the typical ionisation energy (e.g. 3.6 eV in Si). The Bethe-Bloch equation can be modified to suppress the contribution of delta electrons, resulting in the restricted Bethe-Bloch equation, whose scaling at high energies is considerably more mild [176].

The energy loss is an important strategy for charged particle identification, e.g. for the $K^{\pm}$-$\pi^{\pm}$ discrimination scenarios in Section 6.5.1. The dependence of the energy loss on $\beta\gamma$, evident in Figure 7.4a, enables particle identification. Given that particles with the same momentum but different masses have differing $\beta\gamma$, the measurement of the energy loss allows for the determination of a particle's species.

The distribution of the energy loss at a given $\beta\gamma$, typically denoted as the *straggling function*, was derived by Landau [177] and Vavilov [178], and later improved by Bichsel [179] to account for the electron binding energy and a wider range of absorber thickness [180]. Figure 7.4b depicts the straggling functions for a range of silicon thicknesses. The distribution is characterised by a pronounced central peak and a high-energy tail resulting from rare high-energy transfers to atomic electrons. The most-probable value (MPV) of the straggling function is considerably more tractable experimentally than the mean of the distribution, given by Equation 7.13. The measurement of the mean requires capturing the entire energy range, which can be problematic in practice, as the detector response is constrained to a specific energy range, and may become non-linear at high energies. Thus the MPV of the straggling function is used as a practical estimator of the energy loss in tracking detectors.

The energy loss of electrons differs substantially from that of other charged particles, due to differences in kinematics, spin, and their identity [53]. Figure 7.5a depicts the energy loss of electrons in lead as a function of energy. Although the energy loss of electrons is dominated by ionisation at low energies, it is not well described by the Bethe-Bloch equation. Instead, large energy transfers of electrons are described by Moller scattering whereby the impinging electron is scattered by an atomic electron. At increasing energies $E \sim 10$ MeV, energy losses begin to be dominated by the Bremsstrahlung process. Bremsstrahlung describes the emissions of photons as an electron is accelerated in an electric field, in this case that of atomic nuclei of the medium. At energies above 1 GeV typical of particle accelerators, the energy loss of electrons occurs overwhelmingly via Bremsstrahlung.

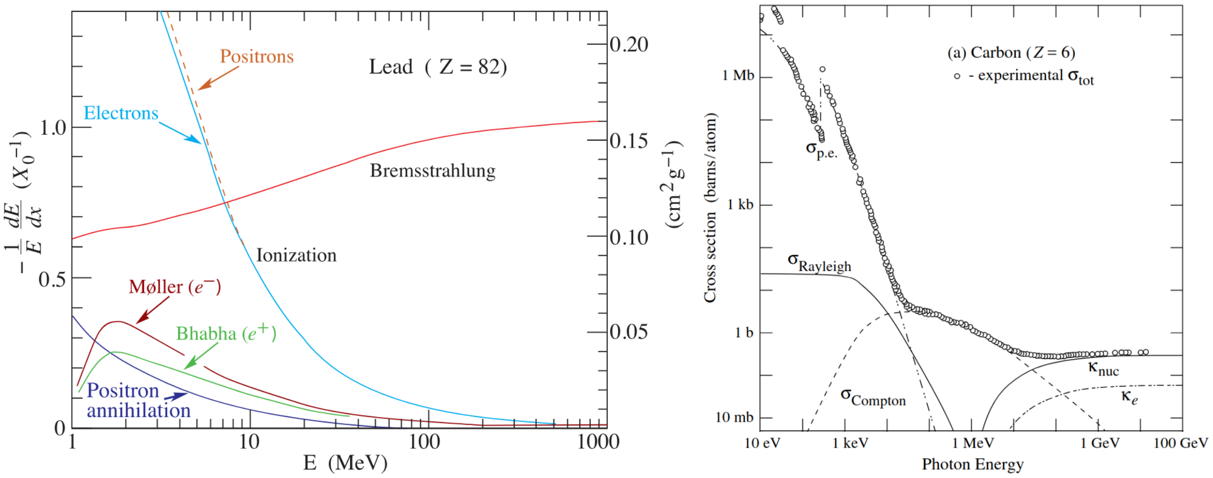


Figure 7.5: Fractional energy loss per radiation length of electrons and positrons in lead (a). Photon cross sections in carbon for a variety of different processes as a function of energy (b). Figures adapted from Ref. [158].

As charged particles traverse matter they undergo a multitude of small-angle scattering processes. Typically, these are dominated by multiple, sequential Coulomb scatterings off of atomic nuclei in a process denoted multiple Coulomb scattering, or *multiple scattering* for short. Multiple scattering can be quantified by the width of the distribution of the scattering angle in the plane perpendicular to the particle trajectory θ_{plane}. In the limit of a large number of scatterings, the scattering angle θ_{plane} is expected to be distributed approximately Gaussian, as per the central limit theorem. Ref. [181] provides an estimate for RMS of θ_{plane} as

$$\theta_{ms} \approx \sqrt{\langle \theta^2_{plane} \rangle} \approx \frac{E_s}{\sqrt{2} pc\beta} \sqrt{\frac{x}{X_0}} \tag{7.14}$$

where $E_s = m_e c^2 \sqrt{\frac{4\pi}{\alpha}}$ is a constant and x/X_0 is the thickness of the medium in radiation lengths. The $1/(p\beta)$ scaling of θ_{ms} highlights that multiple scattering is minimised when particles traverse a medium at high energies. Moreover, multiple scattering can also be reduced by minimising the amount of material that a particle must traverse in a detector, known as the *material budget*. Minimising the material budget in order to reduce the deflection of particle trajectories is a key consideration in tracking detector design.

Photons

The primary mechanisms involved in the interaction of photons with matter are the photoelectric effect, Compton scattering, pair production, and Rayleigh scattering. The cross sections for each of the processes in carbon is depicted on Figure 7.5b. At low energies below $\sim$ 50 keV the photoelectric effect dominates the energy loss of photons in matter. The photoelectric effect denotes the absorption of an energetic photon by an atom, freeing a bound electron and ionising the atom. The energy of the freed electron is given by

$$E_{e^-} = E_\gamma - E_B \tag{7.15}$$

where E_γ denotes the energy of the incoming photon and E_B denotes the binding energy of the electron. In the case of silicon, the freed electron deposits its energy by creating electron-hole pairs as it traverses the medium. A $\sim$ 5.9 keV electron resulting from the photoelectric absorption of a K$_\alpha$ photon from an ^{55}Fe decay will deposit its entire energy almost locally, within $\mathcal{O}(1)$ µm [182]. In addition to the photoelectric effect, Rayleigh scattering also contributes to the interaction of photons at low energies, although with a much smaller cross section. Rayleigh scattering describes the elastic scattering of a photon off an atom. Its contribution to the cross section falls rapidly with energy, as the photon wavelength becomes smaller than the typical radius of an atom.

At higher energies above $\sim$ 50 keV Compton scattering becomes the dominant process, whereby the impinging photon scatters inelastically off a free or shell electron. The photon transfers a fraction of its energy to the electron, and is scattered at an angle.

At even higher energies above $\sim$ 10 MeV, the dominant process is increasingly pair production. Pair production describes the conversion of an energetic photon into an electron-positron pair, as it interacts with the electric fields of atomic nuclei, or atomic electrons, to a lesser degree. Kinematically, the photon energy must be greater than twice the rest mass of the electron, which is evidenced by the $E_\gamma \approx 2m_e c^2 \approx 1.022$ MeV point in Figure 7.5b, above which pair production begins to contribute.

7.2.2 Signal: Induced Current

The formation of a signal in a semiconductor detector begins with the excitation of electron hole pairs by impinging radiation, as described above. The collective, orchestrated motion of charge carriers within the depletion region induces a current at the collection electrode, whose magnitude is given by the Shockley-Ramo theorem [183, 184]

$$i(t) = -q\left(\vec{v} \cdot \vec{E}_Q\right) \tag{7.16}$$

where $q, \vec{v}$ are the charge and velocity of the charge carrier, and $\vec{E}_Q$ is the "weighting field", which relates the carrier motion to the induced current at a given electrode, depending only on geometry [172]. Thus, it is worth emphasising that the formation of a signal begins with the motion of electron hole pairs in the sensitive layer, not necessarily when they arrive at the collection electrode or substrate.

In general, the motion of charge carriers within the depletion region consists of a drift component and a diffusion component. The *drift* of charge carriers within the depletion region due to the electric field is roughly linear according to Equation 7.2, offering a natural avenue to increasing the charge collection speed. Increasing the biasing voltage not only widens the depletion region, guaranteeing that a larger fraction of the deposited charge is collected, but also increases the electric field, particularly after full depletion. The increased charge collection speed has the twin benefits of reducing the likelihood of charge trapping and recombination, which can become significant after radiation damage, thereby increasing the fraction of charge that is collected, but also reduces the rise time of the signal. The drift velocity cannot be increased indefinitely by increasing the electric field, as the linear relationship given in Equation 7.2 breaks down at high electric fields as the drift velocity saturates [53].

In addition, all charge carriers *diffuse* randomly throughout the semiconductor medium as a result of their thermal motion. In this way, what may start as a localised deposition of charge by an impinging particle becomes a charge carrier cloud spread out in a roughly Gaussian profile. The standard deviation of said profile σ_{diff} is given as [173]

$$\sigma_{diff} = \sqrt{2Dt}, \quad D = \mu \frac{k_B T}{e}$$

where D is the diffusion constant and t is the time during which the charge carriers have diffused. Charges generated within the depletion region are unlikely to diffuse very far, since they quickly drift to either the collection electrode or the substrate according to Equation 7.2, minimising the time of their diffusion.

For charges generated outside the depletion region, however, the diffusion component of their motion can be substantial before they reach the depletion region. Indeed, for large enough drift times, or for charge depositions close to pixel boundaries, this could mean diffusing to the depletion region of the neighbouring pixel, and thus inducing a signal in the collection electrode of the neighbouring pixel in a process denoted *charge sharing.* While high charge sharing implies a number of drawbacks related to slow charge collection, charge trapping and recombination, and reduced signal-to-noise ratio (SNR), charge sharing can aid substantially in the reconstruction of a particle hit position by leveraging positional information encoded in the charge of its neighbours. For instance, while a hit close to the pixel boundary may only be reconstructed with a $\sim p/2$ error if the entire charge is collected by the pixel in which the particle impinged, a sharing of half the charge would allow the centre-of-mass to be reconstructed at the pixel boundary, leading to a much lower reconstruction error. In practice, a trade-off must be made in the degree of charge sharing desired in a sensor, with the positional reconstruction accuracy weighed against the radiation hardness, timing performance, occupancy, and resulting hit rate.

7.3 Signal Readout

The induced current introduced in Section 7.2.2 must be integrated in order to obtain a measure of the charge deposited by an impinging particle. Moreover, the resulting signal is generally very small and must be amplified before being processed further. As this is first step following the collection electrode, the amplifier is referred to as a *preamplifier*. The preamplifier may be either DC- or AC-coupled to the collection node, with the latter using a coupling capacitor that allows higher bias voltages at the collection node, albeit at the cost of added input capacitance. Additional steps to increase the SNR, readout speed, or reduce the data rate, are typically performed in subsequent signal processing stages. Since the chips presented in this thesis feature an analogue readout, the reader is referred to Refs. [53, 172] for details on signal processing beyond amplification.

7.3.1 Readout Electronics

The integration of the induced current must be performed at some capacitance, often by a dedicated feedback capacitor of the preamplifier, in which case the preamplifier is said to be *charge-sensitive*. Alternatively, the current may be integrated on the capacitance of the collection node itself, as is done for the chips in this thesis. In other words, the signal at the input of the preamplifier is simply the voltage drop at the collection node due to charge carrier accumulation, and the preamplifier is thus considered *voltage-sensitive*. The observed signal following the preamplifier is a stable baseline defined by the preamplifier working point, which falls sharply as charge from an impinging particle is collected. In order to restore the baseline and maintain the dynamic range of the amplifier, its input must be discharged. The discharge of amplifier input is performed either through some dedicated current (e.g. I_{reset}) or through the leakage current itself, in which case the preamplifier is said to be self-recovering.

While an analogue readout is feasible for the test structures presented in this thesis, for larger detector systems the readout of the full analogue signal can pose a major challenge. Instead, the signal can be *digitised* to a certain precision. There is a tradeoff in the digitisation precision and speed [172], such that for fast applications, only the hit-no hit information is saved [185]. The attainable spatial resolution using the binary method can be computed as the variance of a uniform distribution, integrated from $-p/2$ to $p/2$ corresponding to the pixel boundaries

$$\sigma^2 = \int_{-p/2}^{p/2} x^2 f(x) dx = \frac{1}{p} \int_{-p/2}^{p/2} x^2 dx = \frac{p^2}{12} \tag{7.17}$$

where $f(x)$ is the pdf and p is the pixel pitch. The resolution is then $\sigma = p/\sqrt{12}$, which suffices only for small pitches.

Alternatives to the binary readout method, whereby the collected charge is encoded to higher precision, exist. For instance, the measured pulse height may be passed to comparators, thereby digitising the collected charge. On the other hand, if a reset current discharges the integrating capacitor linearly, then the time required to return to baseline may be exploited to determine the deposited charge, in what is known as the Time-Over-Threshold (ToT) method. It offers a compromise between precision and speed, as the ToT is generally less precise than the full pulse height, but requires fewer bits.

7.3.2 Noise

Noise broadly describes fluctuations of a signal arising from sources unrelated to the underlying measured phenomenon. In the context of semiconductor pixel detectors, the primary noise source in the measurement of an energy deposition is typically the electronic noise resulting from the readout electronics. Minimising electronic noise is one of the key considerations in semiconductor detector design, as noise can quickly degrade the achievable SNR given the relatively low signals generated by ionisation.

Fano Noise

Even if one were to imagine a detector with negligible electronic noise, a measurement of the energy deposited by a particle would still be smeared by statistical fluctuations in the number of generated electron hole pairs. The number of electron hole pairs N_{eh} generated by an energy deposition has the mean $\langle N_{eh} \rangle = E/E_i$, where E_i denotes the ionisation energy. Assuming a Poisson distribution would imply a standard deviation of $\sigma_{N_{eh}} = \sqrt{E/E_i}$, which would give a measurement resolution of $\sigma_{N_{eh}}/N_{eh} = \sqrt{E_i/E}$. In practice, however, the observed standard deviation is significantly smaller due to the contributions from other excitation mechanisms such as phonons [172], though a complete understanding does not yet exist [173]. The deviation from the expected standard deviation from statistics alone is given by the *Fano factor*

$$\frac{\sigma_{N_{eh}}}{N_{eh}} = \sqrt{F \cdot \frac{E_i}{E}} \tag{7.18}$$

The Fano factor for silicon is $F \approx 0.1$, setting the intrinsic energy resolution that can be achieved in a silicon detector.

Fixed Pattern Noise

Measurements that combine the signal of multiple pixels, such as reconstructing a particle track position from the signal of neighbouring pixels, are subject to an additional type of noise resulting from variations between them. *Fixed Pattern Noise (FPN)* denotes the noise resulting from the differing response of neighbouring pixels due to gain and baseline differences introduced by manufacturing variations. FPN is characterised by varying spatially, for instance across a pixel matrix, but being temporally constant. Calibration measurements can be used to scale the response of each pixel, minimising FPN, as is done in Section 9.6.3 for the chips studied in this thesis.

Electronic Noise

Electronic noise results from statistical fluctuations in the readout electronics. The primary noise sources in silicon detectors include thermal noise, shot noise, and $1/f$ noise. *Thermal noise*, or Johnson noise, describes voltage fluctuations resulting from the random thermal motion of charge carriers in a resistive circuit element. The spectral density up to high frequencies is constant, and proportional to the temperature of the medium.

Shot noise describes fluctuations in the number of charge carriers overcoming some potential barrier. In semiconductor detectors the primary source of shot noise is the stochasticity of thermal electron-hole pair generation in the *pn*-junction, leading to fluctuations in the leakage current. The spectral density is likewise constant, and proportional to the leakage current.

Flicker noise, or $1/f$ *noise*, encompasses noise whose spectral density scales as $1/f^{\alpha}$ for some $\alpha = 0.5, ..., 2$ [53], and thus dominates at low frequencies. In electronics $1/f$ noise is associated with the trapping of charge carriers. The correlated double sampling technique briefly introduced in Section 8.4 can be used to mitigate $1/f$ noise [186].

The contribution of noise to the measured signal is often quantified by measuring the equivalent noise charge (ENC). The *equivalent noise charge* represents the fluctuation in input charge (i.e. N_{eh}) that would correspond to the observed noise in the output signal (e.g. $\Delta\mathrm{V}$). The ENC is obtained by scaling the measured noise by the charge-to-voltage gain of the detector

$$ENC = \frac{\Delta\mathrm{V}}{G_Q} \tag{7.19}$$

where G_Q is the charge-to-voltage gain of the detector[1] under consideration. In this way, the contribution of all noise sources can be compared, regardless of where in the signal processing chain they appear. Moreover, by expressing the noise in terms of input charge, and thus as a fraction of the signal, the observed noise can be meaningfully compared across detectors as its magnitude no longer depends on the amplifier gain.

7.4 Monolithic Pixel Detectors

Semiconductor detectors function by converting the energy deposited by an impinging particle into an electrical signal. Typically the charge formation and collection described in Section 7.2 is performed on a high-resistivity piece of silicon where the depletion region can extend over a large area of the sensor. The intricate readout electronics, on the other hand, require a low-resistivity piece of silicon, where the signal can be amplified and digitised. In the hybrid approach the sensor and the readout electronics are fabricated on two separate silicon dies and connected via a micro-connection, typically a bump bond. In this way, the readout chip and the sensor can be optimised independently.

MAPS Overview

Monolithic Active Pixel Sensors (MAPS) combine the sensor and readout chip onto the same silicon die. Integrating sensing and readout electronics onto the same silicon die introduces challenges, including isolation of the transistors. In order to ensure proper functioning of PMOS transistors, they must be fabricated within an n-doped well. The n-well must, in turn, be isolated from the active layer where electron-hole pairs are generated, since it would otherwise compete with the n-type collection electrode for electrons. Two strategies for achieving this are either the use of a large collection electrode into which the readout circuitry is placed [187], or separate deep p-wells which ensure isolation from the active sensing layer, as depicted Figure 7.6. The latter approach is adopted for the chips presented in this thesis, which entails the benefit that the n-well collection electrode can be miniaturised, reducing the junction capacitance, and thereby benefitting the SNR.

MAPS offer a variety of advantages with respect to their hybrid counterparts, including a

---

[1] For the CE-65 family of chips studied in this thesis, G_Q corresponds to $\gamma_{\mathrm{ccf}}^{-1}$ defined in Equation 9.5.

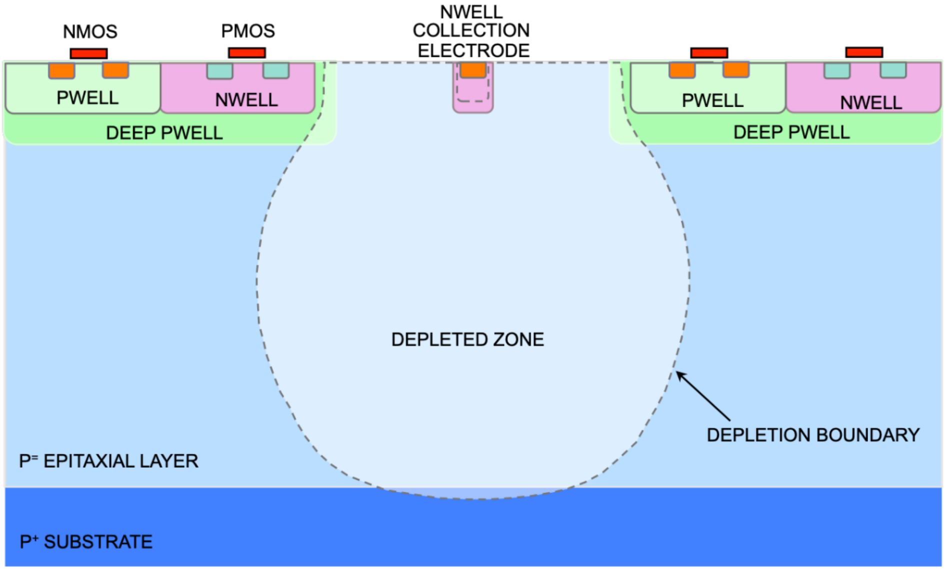


Figure 7.6: Schematic of a MAPS pixel cross section. Adapted from Ref. [46].

lower material budget, reduced power consumption, and smaller pitch/miniaturisation, as bump bonding is not needed. Nevertheless, the high rate and high radiation environment inherent to hadron colliders has meant that MAPS have seen limited adoption at hadron colliders. These conditions can prove challenging for MAPS, where the limited depletion region means charge collection often contains a diffusive element which can lead to much slower charge collection times, and make charge carriers more prone to charge trapping and recombination due to radiation defects. Hybrid sensors, on the other hand, are more well-suited due to the ability to apply large voltage biases to the high-resistivity sensor, and thereby lead to the depletion of large areas of the chip. The resulting drift-dominated charge collection is considerably faster, and more radiation tolerant.

Moreover, while the use of existing commercial CMOS technologies greatly reduces cost, typical CMOS technologies include a very shallow $\mathcal{O}(10)$ µm high-resistivity layer on the silicon substrate. This shallow layer, termed the *epitaxial* layer, acts as the sensing layer in MAPS sensors, as depicted on Figure 7.6. Most of the silicon substrate thus does not partake in charge generation and collection, acting mostly as mechanical support [172]. The thin sensing layer entails a small charge deposition, as described in Section 7.2. The small generated signal thus places stringent constraints on the tolerable noise level, and impacts the achievable energy resolution.

Applications

The radiation background in future lepton colliders will be significantly lower than in hadron colliders, removing a major limitation for MAPS-based detectors. With the push for high-precision vertex detectors at future lepton colliders, MAPS offer a promising avenue due to their low material budget, which reduces multiple scattering and improves spatial resolution. A number of (proposed) future collider experiments are actively exploring MAPS for vertexing, including the FCC-ee [47, 188], CLIC [189], ILC [190], and CEPC [191]. In particular, MAPS have been proposed as the enabling technology for the high-precision vertex detector for all 4 detector concepts introduced in Section 3.5.

MAPS have already seen adoption at heavy-ion experiments, including ALICE at the LHC

[192, 193] and STAR at RHIC [194], where radiation environments are considerably more mild than in proton-proton collisions. The synergy between ongoing research efforts on MAPS development for heavy-ion experiments and future lepton colliders is exemplified by the latest upgrade of ALICE's Inner Tracking System (ITS), dubbed the ITS3 [46]. The ITS3 upgrade plans to replace the ITS2's [192] three innermost layers with fully cylindrical, wafer-scale bent sensors, targeting an excellent ~ 5 µm spatial resolution and an ultra-low material budget of 0.07% X_0 per layer. Refs. [195, 196] highlight that the FCC-ee vertex detector requirements are generally comparable to those of the ALICE ITS3 upgrade, with requirements on the spatial resolution being somewhat more stringent, while baseline material budget constraints are more relaxed[2]. The details of a MAPS test structure targetting both the ALICE ITS3 upgrade and future colliders such as the FCC-ee are presented in the following chapter, before delving into its characterisation thereafter.

---

[2] Ref. [47] explores the benefits of a reduced material budget at the FCC-ee.

8 CE-65 Chip

In December 2020 the first set of small-scale test structures targeting the ALICE ITS3 upgrade [46] were produced using the TPSCo 65 nm CMOS imaging process. This Multi-Reticle Layer Run 1 (MLR1) submission targeted the systematic exploration and validation of key aspects of the 65 nm technology, including process parameters, in-pixel electronics, and front-end designs, in preparation for the wafer-scale MAPS sensors of ITS3. The MLR1 submission featured a number of distinct test-structures targeting specific challenges of the ITS3 upgrade, three of which are listed here:

- Analogue Pixel Test Structure (APTS) [197]: A small 6×6 pixel matrix, of which the inner 4×4 are active, with direct analogue readout, optimized for studying charge collection efficiency, noise performance, and collection electrode geometry variations
- Digital Pixel Test Structure (DPTS) [198]: A larger 32×32 pixel matrix with time-encoded digital readout, developed to evaluate pixel-level digital electronics
- Circuit Exploratoire 65 nm (CE-65) [199]: A larger test structure featuring a 64×32 pixel matrix combined with analogue readout, designed to investigate scalability, and process properties in the 65 nm technology

The MLR1 prototypes provided critical insights into the feasibility of 65 nm MAPS for ITS3, validating key process parameters [4, 196–204].

While the subsequent ER1 submission targeted primarily the ALICE ITS3 upgrade through the fabrication of the MOSS and MOST sensors, significant room on the wafer was devoted to the advancing of earlier test-structures, including the CE-65v2: the evolution of CE-65.

8.1 Chip Details

The CE-65 and its evolution, the CE-65v2, were developed to investigate the charge collection and electrical properties of the 65 nm CMOS process, targetting not only the ITS3 upgrade, but future applications of MAPS, including the FCC-ee. Both the CE-65v1 and the CE-65v2 feature a large matrix read out in an analogue way, as depicted on Figures 8.1a and 8.1b.

While the CE-65v1 implements three separate in-pixel amplification schemes: an AC-coupled preamplifier (AC), a DC-coupled preamplifier (DC), and a DC-coupled source-follower (SF), the CE-65v2 features only the AC-coupled pre-amplifier. The CE-65v1 chip is split into three submatrices, each corresponding to one of the amplification schemes. The CE-65v2 expands on the CE-65v1 by including a wide variety of chip variants targeting additionally the exploration of process and pitch variations, and introduced the staggered pixel arrangement as a way to gauge the importance of matrix geometry on the achievable single point resolution. The four *exploration axes* that are targeted by the CE-65 family of test structures are listed below:

- Amplification Scheme: AC-amp., DC-amp., SF
- Process variation: Standard, Modified, Modified w/ Gap

- Pitch variation: 15 µm, 18 µm, 22.5 µm
- Matrix geometry: square vs staggered

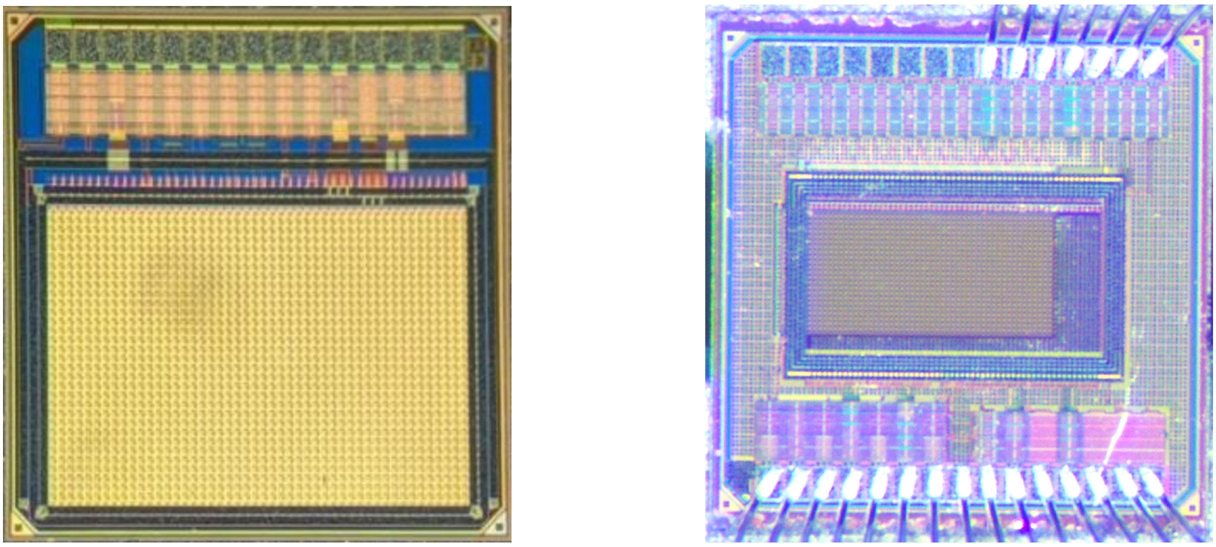


Figure 8.1: Photographs of the CE-65v1 chip (a), taken from Ref. [46], and CE-65v2 chip (b), provided by IPHC Strasbourg. The chip dimensions are 1.41 mm × 1.41 mm, and 1.5 mm × 1.5 mm, respectively.

The adopted amplification scheme has ramifications on the noise performance of the chip and thus on the measured energy resolution. The AC-coupled preamplifier was selected as the baseline in CE-65v2 after demonstrating superior signal-to-noise characteristics in CE-65v1, as shown in Chapter 9.

Adopting a reduced pixel pitch allows for an improvement in the spatial resolution, but comes with notable drawbacks in terms of power consumption, readout-rate, as well as challenges during manufacturing. Thus it becomes necessary to explore the pitch size at which the desired spatial resolution can be achieved, without compromising other aspects of the chip and overall system design.

The process variations are detailed in Section 8.2. The variations are centered around the inclusion of a a deep low-dose n-type implant between the epitaxial layer and the CMOS circuitry [205], enabling a more uniform formation of the depletion region in the epitaxial layer. These result in different levels of charge sharing, directly affecting the achievable resolution for each of the sensor variants.

The final exploration axis is the inclusion of a staggered matrix geometry where every second column is staggered by half the pixel pitch, as introduced in Section 8.5. This, in principle, reduces the number of neighbours that a seed pixel would have in a cluster, and thus promises an improved resolution at higher thresholds.

8.2 Process Modifications

Process variations following the lessons learned during the development of the ALPIDE [185] chip, targeting the ITS2 upgrade, resulted in considerable improvements in radiation hardness and charge collection speed.

8.2.1 Standard Process

Figure 8.2a illustrates the cross section of a pixel produced in the *Standard* process, consisting of an n-well collection electrode, and in-pixel CMOS circuitry that is isolated from the epitaxial layer by a deep p-well. The deep p-well provides a charge-reflective boundary between the epitaxial layer and the circuitry, preventing the n-wells housing the PMOS transistors from competing with the collection electrode for the generated electrons. The depletion region begins to develop at the n-well collection electrode and follows a balloon shape in the eptaxial layer, as the reverse biasing voltage is increased. This proceeds until the p+ substrate is reached, at which point the strong doping gradient inhibits further depletion. Due to the limited depth of the epitaxial region of $\mathcal{O}(10\,\mu\mathrm{m})$, the lateral region remains largely undepleted, resulting in diffusion-dominated charge collection. In effect, charge collection is slow and subject to charge trapping, whilst exhibiting substantial charge sharing between pixels.

8.2.2 Modified Process

In order to address the shortcomings of the Standard process, a planar junction spanning the pixel width was introduced [205]. Figure 8.2b depicts the *Modified* process which has, in addition, a deep low-dose n-type implant between the epitaxial layer and the CMOS circuitry. The development of the depletion region now begins at the interface of the n-type implant and the epitaxial layer. Even at low bias voltages, the depletion region spans the width of the pixel and begins to extend vertically. While this ensures full depletion of the epitaxial layer, it also entails a substantial increase in the sensor capacitance as the region around the n-well collection electrode remains undepleted [205]. This is addressed by providing higher biasing voltages, which then deplete the region in the n-type implant between the n-well collection electrode and the deep p-wells. At a high biasing voltage the capacitance is defined by the geometry of the n-well collection electrode itself. Indeed, during the laboratory characterisation of the CE-65 family of chips detailed in Chapter 9, capacitance was found to vary little across process variations at full depletion. While a larger biasing voltage is needed to operate chips at low sensor capacitance than in the Standard process, the lateral depletion of the Modified process allows for the development of the electric field. The electric field induces drift-dominated charge collection, resulting in faster charge collection and reduced charge sharing. The resulting improved radiation hardness was one of the primary motivations for the introduction of the Modified process [46].

8.2.3 Modified with Gap Process

Nevertheless, the Modified process displayed considerable drops in efficiency towards the pixel edges at mild irradiation doses of 10^{14} 1 MeV n_{eq} cm^{-2}. Figure 8.2c depicts the *Modified with Gap* process, which adopts the same layout as the Modified process, but includes gaps at the pixel edges [206]. The gaps enable the development of the electric field at the edges, resulting in even less charge sharing than in the Modified process.

While the process modifications were initially developed for the 180 nm TowerJazz technology, the adoption of the TPSCo 65 nm technology entails a reduction in the depth of the epitaxial layer by over a factor 2 [46]. Accordingly, the incomplete lateral depletion of the Standard process is magnified, with an even larger fraction of the epitaxial layer remaining undepleted, for equivalent pixel pitches. Thus, the impact of the process modifications is amplified with respect to the 180 nm process.

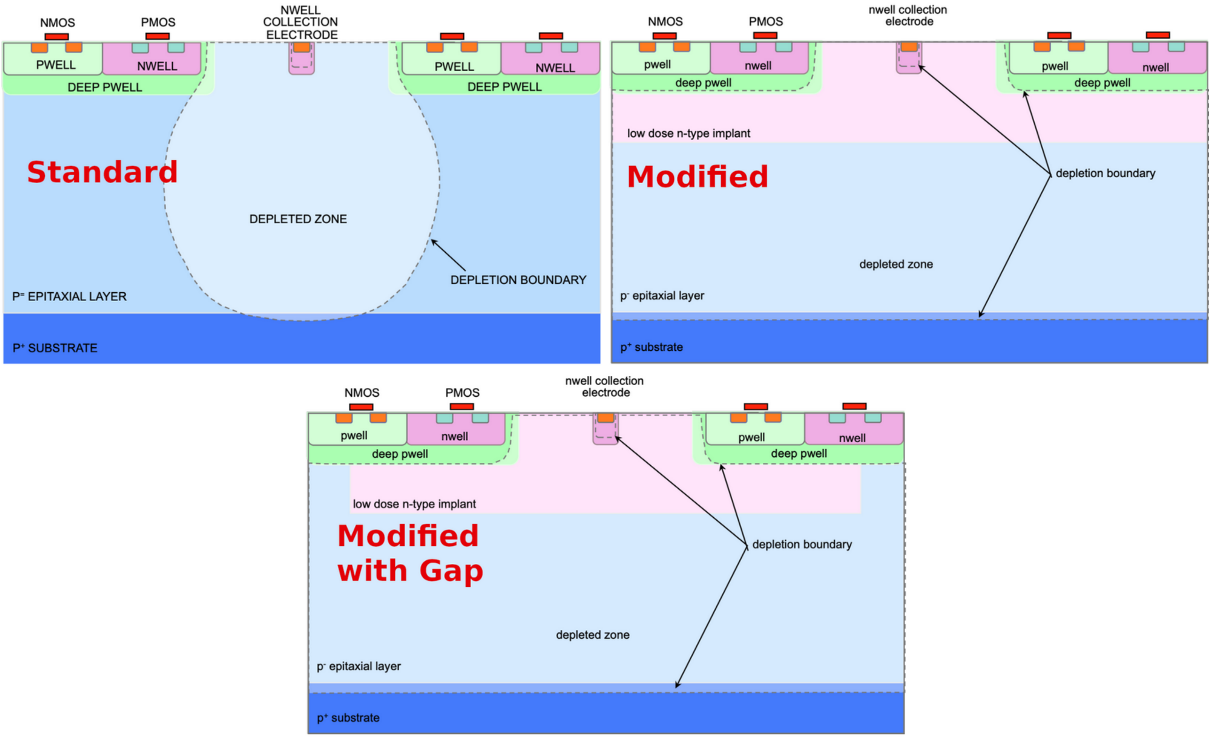

NMOS
PMOS
NWELL COLLECTION ELECTRODE
PWELL
NWELL
DEEP PWELL
Standard
DEPLETED ZONE
DEPLETION BOUNDARY
P$^-$ EPITAXIAL LAYER
P$^+$ SUBSTRATE
NMOS
PMOS
nwell collection electrode
pwell
nwell
deep pwell
low dose n-type implant
Modified
depletion boundary
depleted zone
p$^-$ epitaxial layer
p$^+$ substrate
NMOS
PMOS
nwell collection electrode
pwell
nwell
deep pwell
low dose n-type implant
Modified with Gap
depletion boundary
depleted zone
p$^-$ epitaxial layer
p$^+$ substrate

Figure 8.2: Schematic of the CE-65(v2) pixel cross section for the Standard, Modified, and Modified with Gap, process modifications. Adapted from Ref. [46].

8.3 Amplification Scheme

The adoption of in-pixel gain amplifiers is one of the key components in the processing of an analog signal in a silicon detector. The signal amplification substantially improves the signal-to-noise ratio. The design of the in-pixel amplifier involves a trade-off between gain, speed, noise, and power consumption [207]. One of the key objectives of the CE-65v1 test structure is studying the impact that in-pixel signal amplification scheme has on the achievable spatial resolution. As mentioned in Section 8.1, the CE-65v2 adopted the AC-coupled preamplifier only, and thus the discussion in this section pertains primarily to the CE-65v1 chip.

8.3.1 Source-Follower

The three submatrices of the CE-65v1 chip are depicted on Figure 8.3a, with an equivalent number of rows for all four variants. Figure 8.3b depicts the in-pixel circuitry of the rightmost submatrix of the CE-65v1 chip, consisting of a simple 3-T source-follower (SF). The circuit acts as both a source-follower and a level-shifter to the 1.2 V domain. The current source is implemented as a NMOS transistor, below M1. The collection node is biased by a resetting diode at a voltage of V_{reset}. Since the bias is DC coupled with the gate of M1, V_{reset} cannot exceed the supply voltage at the drain of M1, given by $V_{DD33} = 3.3$ V.

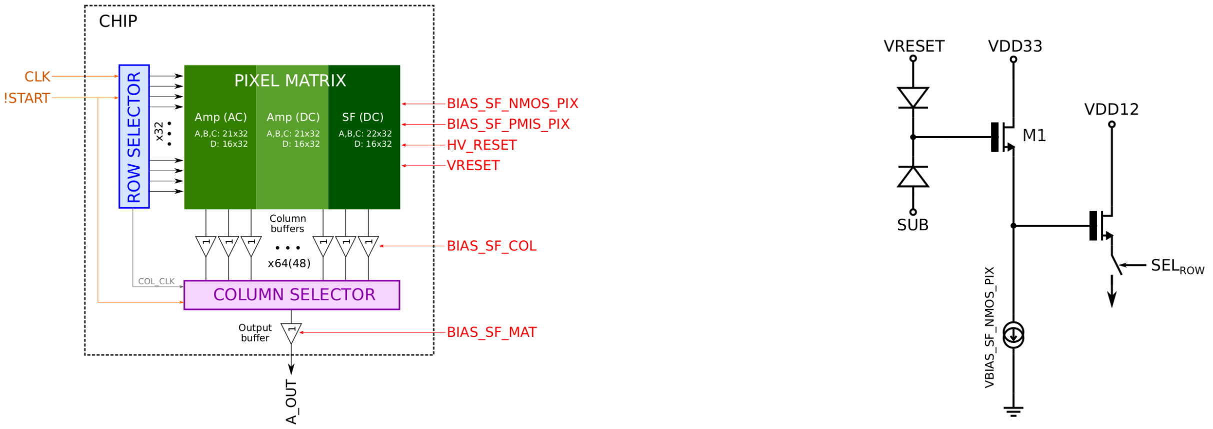

CHIP
CLK
!START
ROW SELECTOR
x32
PIXEL MATRIX
Amp (AC)
A,B,C: 21x32
D: 16x32
Amp (DC)
A,B,C: 21x32
D: 16x32
SF (DC)
A,B,C: 22x32
D: 16x32
BIAS_SF_NMOS_PIX
BIAS_SF_PMIS_PIX
HV_RESET
VRESET
Column buffers
x64(48)
BIAS_SF_COL
COL_CLK
COLUMN SELECTOR
Output buffer
BIAS_SF_MAT
A_OUT
VRESET
VDD33
VDD12
M1
SUB
SEL$_{ROW}$
VBIAS_SF_NMOS_PIX

Figure 8.3: Diagram of the three submatrices of the CE-65v1 chip, and its readout (a). Biases are marked in red. Schematic of the CE-65(v2) in-pixel source-follower amplification scheme (b). Figures taken from Ref. [208].

8.3.2 DC Preamplifier

The second amplification scheme at the centre submatrix is the DC-coupled preamplifier, depicted on Figure 8.4a. It implements a more sophisticated architecture consisting of a self-biased amplifier and a pmos source-follower acting as a level shifter to the 1.2 V domain. The sensing node is connected directly to the gate of the input transistor, meaning no resetting diode can be implemented. Instead, the collection electrode is maintained at the operating voltage of the pre-amplifier, typically around 1 V. In this way a gain of $\sim$ 5 times that of source-follower submatrix is obtained.

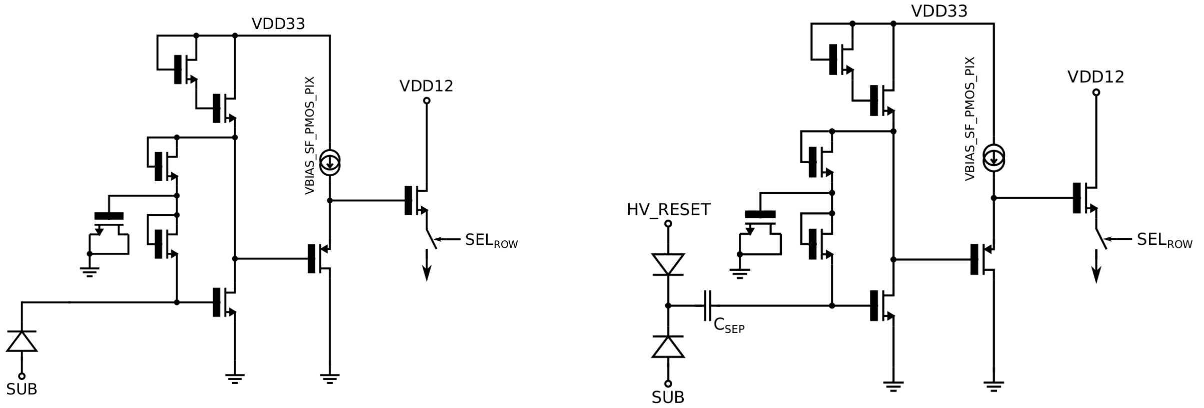

VDD33
VBIAS_SF_PMOS_PIX
VDD12
SEL$_{ROW}$
SUB
VDD33
VBIAS_SF_PMOS_PIX
VDD12
HV_RESET
SEL$_{ROW}$
C$_{SEP}$
SUB

Figure 8.4: Schematic of the CE-65(v2) in-pixel DC (a) and AC (b) amplification schemes. Figures taken from Ref. [208].

8.3.3 AC Preamplifier

Figure 8.4b depicts the final amplification scheme: the AC-coupled preamplifier. The same in-pixel electronics as for the DC-coupled preamplifier are adopted, with the important distinction that the sensing node is separated from the input of the readout stage by a capacitance of $\sim$ 10 fF [199]. While this comes at the cost of a reduced gain with respect to the DC variant due to parasitics introduced by the capacitor, this allows the application of much larger biasing voltages at the sensing node via a resetting diode, similar to the SF amplification scheme. Throughout this thesis, a baseline value of $HV_{reset} = 10$ V was chosen, as it was found to fully deplete the chip. The three amplification schemes can thus be ordered with respect to gain as SF $<$ AC $<$ DC, and with respect to the depletion of the epitaxial layer as DC $<$ SF $<$ AC.

8.4 Chip Readout

The rolling shutter readout of the CE-65 family of chips serialises the output from every pixel into a contiguous structure consisting of the analog signals, read in a column-major fashion. It is implemented similarly for both versions of the CE-65 chip, and thus the readout of the CE-65v1 is detailed. Figure 8.5 depicts the readout of the CE-65v1 chip.

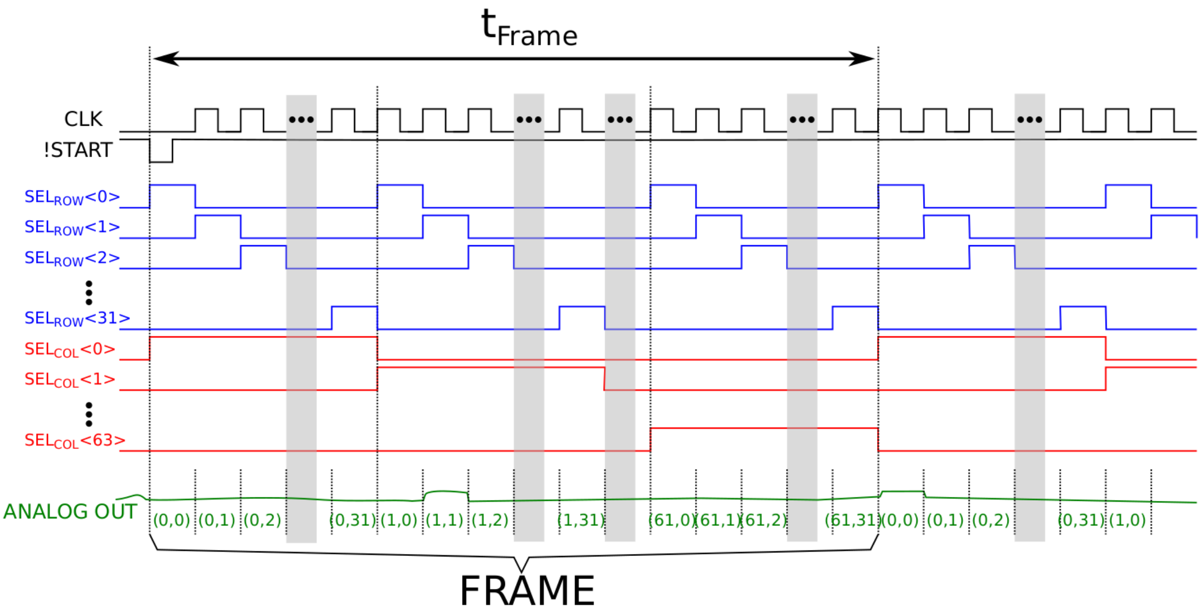


Figure 8.5: Readout of a single frame of the CE-65v1 chip. CLK and !START signals (black) are provided externally, while row (blue) and column (red) selector signals are generated internally. Figure taken from Ref. [208].

Readout begins with an externally provided start signal (!START), which drives the analog signal of pixel (0,0) at the corner of the chip to the output. A clock signal (CLK) defining the readout period for each pixel is provided. While the CLK can be operated at a frequency of up to 40 MHz [199], throughout this thesis the presented measurements were performed at 10 MHz. The rising edge of each CLK signal defines two sets of internal signals which select the column and row corresponding to the pixel that is to be read out, respectively. After all pixels of the matrix have been read out, the subsequent clock cycle again initiates the read out of pixel (0,0), defining the new frame.

The pixel signal is shifted away from the baseline as charges generated in the epitaxial layer generate a current at the electrode following the Shockley-Ramo theorem, introduced in Section 7.2.2. This is depicted for a pixel of the SF submatrix on Figure 8.6, where the negative polarity of the signal can be seen, before a return to baseline. For the AC and DC submatrices the inverting amplifier flips the polarity.

In order to minimise noise contributions arising from low-frequency sources and offsets related to FPN, Correlated Double Sampling (CDS) is employed, whereby time-adjacent samples are

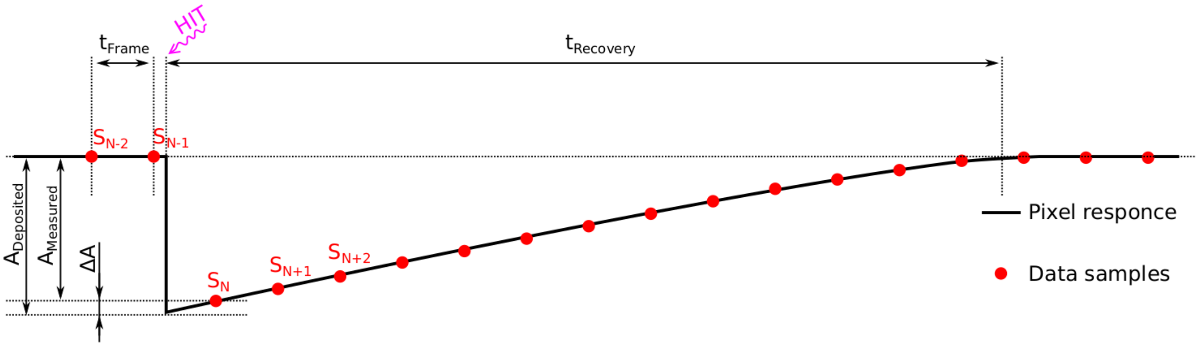


Figure 8.6: Illustration of the pixel response to deposited charge for a source-follower pixel. Figure taken from Ref. [208].

subtracted:

$$A_{meas.} = S_N - S_{N-1} \tag{8.1}$$

where S_N is the signal measured for a given pixel during the N^{th} iteration of the readout, and $A_{meas.}$ is the measured amplitude which is proportional to the generated charge in the pixel.

8.5 Matrix Arrangement

The FASTPIX [209] prototype chip implements hexagonal pixels with pitches ranging from 8.66 to 20 µm in the TowerJazz 180 nm process. The hexagonal geometry minimises the maximal distance from the pixel edges to the collection electrode. While in a square pixel this is given by $p/\sqrt{2}$, where p is the pixel pitch, in a hexagonal pixel with equal area this distance is reduced to $0.88 \cdot p/\sqrt{2}$ [209]. The smaller distance between pixel edges and the collection electrode, coupled with the decreased number of neighbours, results in considerably reduced charge sharing. The hexagonal pixel geometry thus constrains charge to typically be shared by 3 pixels or less, offering a variety of benefits associated with a higher charge per pixel.

In an effort to emulate the benefits of the hexagonal pixel geometry without adopting truly hexagonal pixels, the CE-65v2 chip was implemented in a staggered matrix arrangement, in addition to the square arrangement.

The *square arrangement*, depicted on Figure 8.7a, defines the baseline design of the chip, and was adopted for all other exploration axes of the CE-65v2 test structure. In effect, this means there 9 variants corresponding to the 3 process modifications and 3 pixel pitches.

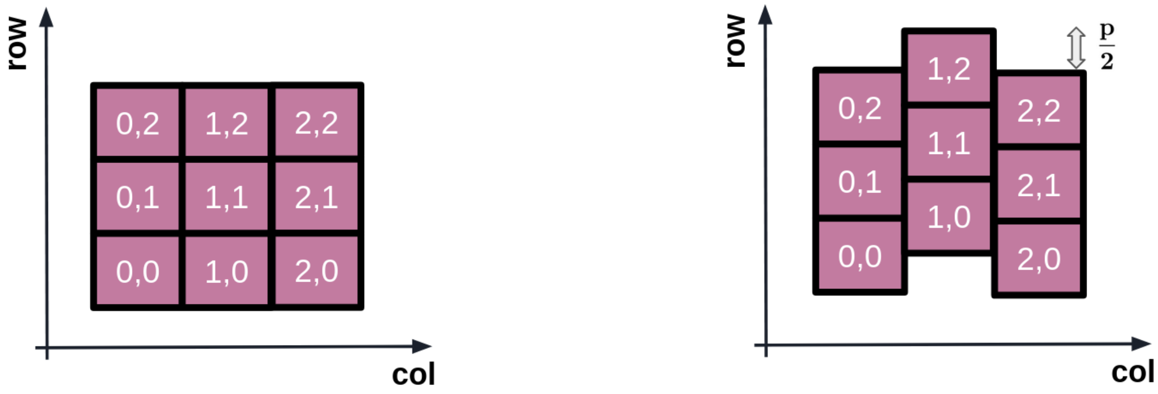


Figure 8.7: Illustration of the square (a) and staggered (b) matrix arrangements adopted in the CE-65v2 chip.

The *staggered arrangement*, depicted on Figure 8.7b, was adopted only for the 22.5 µm and 18 µm pixel pitches, and thus for 6 variants in total. In the staggered arrangement every second column is shifted by $p/2$ vertically. The shift results in a reduction in the immediate number of neighbours from 8 in the square arrangement to 6. While the collection electrodes of the neighbouring pixels now follow a hexagon, it is important to distinguish the arrangement of the neighbours from the shape of the pixel boundaries, which remain squares. This entails that unlike in the FASTPIX chip where the charge collected by the seed pixel was observed to be substantially higher than in the square arrangement, the staggered arrangement shows little difference with respect to the square arrangement. However, the reduction in the number of neighbours results in an increase in the number of clusters consisting of 3 pixels or less, as demonstrated in Section 10.2.6. This reduction, coupled with a geometry favourable for cluster position reconstruction, resulted in sizeable improvements to the spatial resolution of the staggered arrangement variants.

8.6 Summary of V1 and V2 Design Differences

A summary of the properties and differences between between the CE-65v1 chip and the CE-65v2 is listed below:

Table 8.1: Selected design parameters of CE-65v1 and CE-65v2. Parameters only present in the 22.5 µm Standard process CE-65v1 variant are written in parentheses.

Parameter	**CE-65v1**	**CE-65v2**
Sensor dimensions	64×32 pixels (48×32 pixels)	48×24 pixels
Process modifications	Standard, Modified, Gap	
Readout	Rolling shutter	
Pitch	15 µm (25 µm)	15, 18, 22.5 µm
Amplification scheme	AC-amp., DC-amp., SF	AC-amp.
Matrix geometry	Square	Square, Hexagonal/Staggered

9 Laboratory Characterisation

The characterisation of novel MAPS technologies begins in a controlled laboratory setting, where the basic functionality of the chip is verified. Fundamental properties including charge collection and noise performance are evaluated. For the CE-65 family of sensors, this initial evaluation focused on critical performance metrics revolving around the sensor's response to X-rays and noise characteristics. Using a ^{55}Fe source, the laboratory studies detailed in this chapter probed the charge collection properties of the CE-65v1 and CE-65v2 variants, including the depletion of the epitaxial layer, charge sharing, collection efficiency, and gain uniformity across the pixel matrix. Complementary noise performance studies were conducted to observe both the pixel baseline and ENC, providing a picture of some of the sensors' limitations.

The laboratory characterization served as a crucial first step in investigating three of the four *exploration axes* introduced in Section 8.1: amplification scheme, process modification, and pitch variation. It served as a precursor to subsequent testbeam studies, where additional performance aspects, including the spatial resolution and detection efficiency, were evaluated to demonstrate the full potential of sensors fabricated in the 65 nm TPSCo CMOS process. This chapter begins by detailing the data acquisition (DAQ) setup before delving into the analysis strategy and clusterisation procedure. The gain calibration of the CE-65v1 and CE-65v2 chips is presented, emphasizing key performance parameters such as charge collection efficiency and gain uniformity. Finally, key trends and results are summarized.

> *Parts of the ^{55}Fe source characterisation covered in this chapter have been published in Ref. [4]. Similar phrasing was used for relevant sections.*

9.1 Lab Setup

9.1.1 Data Acquisition Setup

The DAQ setup at the University of Zurich is depicted on Figure 9.1. The custom test setup was developed by INFN Cagliari in collaboration with CERN targeting all of the MLR1 test structures introduced in Chapter 8. The setup consists of a DAQ board, a proximity board, and a carrier board housing the CE-65 chip, all interconnected via PCIe connectors. Various connectors can be seen which power the test setup, provide voltage biases, and read out data. A brief overview of the core components is given below, but the interested reader is encouraged to consult Ref. [210] for details.

At its core Altera Cyclone IV-based DAQ board serves to read out and format the chip data, transferring it to the connected PC via the USB interface visible at the left edge of Figure 9.1, where it is further processed using the DAQ software framework described in Section 9.1.2. The DAQ board includes a configurable clock (up to 40 MHz), enabling tests of different CE-65 readout speeds. In the presented results, the clock speed was limited to 10 MHz. In addition, reverse biasing voltages are applied to the CE-65 chip via the LEMO connectors visible on top of the DAQ board. Power is supplied externally to the DAQ board via the green 3-pin screw

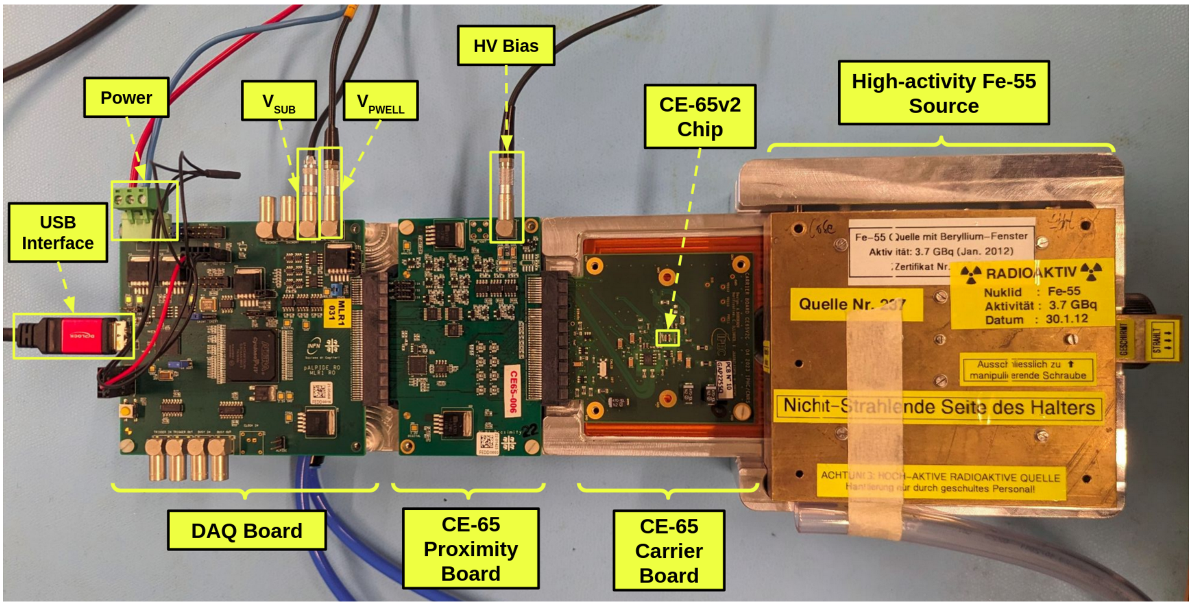

HV Bias
Power
V_{SUB}
V_{PWELL}
CE-65v2 Chip
High-activity Fe-55 Source
USB Interface
Fe–55 Quelle mit Beryllium–Fenster
Aktivität: 3.7 GBq (Jan. 2012)
Zertifikat Nr.
RADIOAKTIV
Nuklid : Fe-55
Aktivität : 3.7 GBq
Datum : 30.1.12
Quelle Nr. 237
Ausschliesslich zu manipulierende Schraube
Nicht–Strahlende Seite des Halters
ACHTUNG: HOCH–AKTIVE RADIOAKTIVE QUELLE
Handhabung nur durch geschultes Personal!
DAQ Board
CE-65 Proximity Board
CE-65 Carrier Board

Figure 9.1: Laboratory test setup of the CE-65v1 and CE-65v2 chips at the University of Zurich.

terminal connector. The pins to the right of the power connector are used to control which of the MLR1 prototypes is being tested.

Each of the proxy boards is test structure specific, and houses DACs providing voltage and current biasing, as well as ADCs digitizing the output for transfer to the DAQ board. For CE-65, readout is performed using a 16-bit ADC.

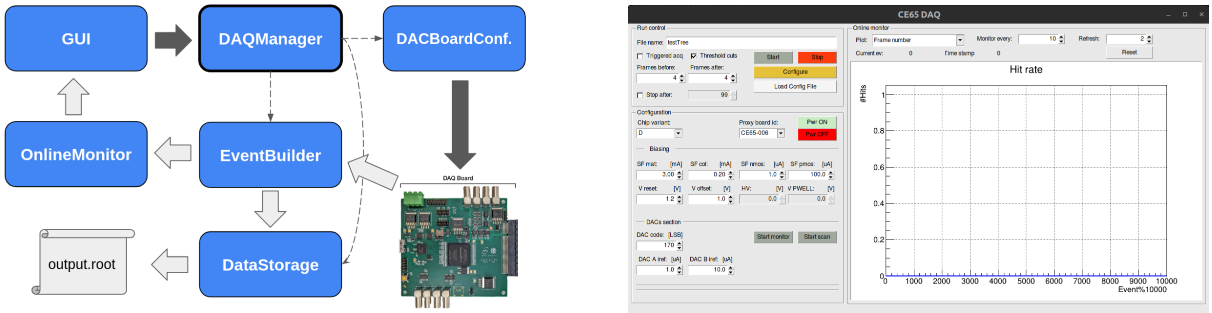
The test system was mounted on an Aluminium structure hosting the DAQ board, the proxy board, and the carrier board. Temperature control was provided using the Huber Minichiller to circulate water below the Aluminium structure. Dry air was used to prevent condensation. A Rhode & Schwarz HMP2030 power supply was used to power the DAQ board and provide V_{PWELL}, V_{SUB}, and high-voltage (HV) bias voltages to the CE-65 chip. An RC-filter was used to mitigate high-frequency noise from the power supply.

The ^{55}Fe source holder was slid above the chip before data taking leaving a distance of a couple cm from the source to the chip. The activity of the source was estimated at $\sim$ 200 MBq at the time of data taking.

9.1.2 Data Acquisition Code

The DAQ code consists of a C++-based Graphical User Interface (GUI) implemented using the ROOT framework, initially developed by the IPHC Strasbourg. The framework consists of modules serving separate functions in the data processing chain. Figure 9.2a depicts a high-level schematic containing the core modules. The GUI, depicted on Figure 9.2b, serves as the direct point of interaction for the user. User-chosen configurations, including the chip variant or proxy board ID, are specified in the GUI. Data taking commands including the starting and stopping of runs are also undertaken in the GUI.

The GUI passes user commands to the DAQManager, which serves as the core steering module of the framework. Prior to data taking a run must be *configured*, whereby biasing conditions are chosen and passed on to the DACBoardConfigurator. The DACBoardConfigurator is responsible for direct interaction with the physical DAQ board introduced in Section 9.1.1. This includes low-level hardware interactions such as the powering of the chip, along with setting of DAC


GUI
DAQManager
DACBoardConf.
OnlineMonitor
EventBuilder
DataStorage
output.root
DAQ Board
CE65 DAQ
Run control
Start
Stop
Configure
Load Config File
Configuration
Pwr ON
Pwr OFF
Biasing
DACs section
Start monitor
Start scan
Online monitor
Reset
Hit rate
#Hits
1
0.8
0.6
0.4
0.2
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Event%10000

File name: testTree
☐ Triggered acq ☑ Threshold cuts
Frames before: 4
Frames after: 4
☐ Stop after: 99
Chip variant: D
Proxy board id: CE65-006
SF mat: [mA] 3.00
SF col: [mA] 0.20
SF nmos: [uA] 1.0
SF pmos: [uA] 100.0
V reset: [V] 1.2
V offset: [V] 1.0
HV: [V] 0.0
V PWELL: [V] 0.0
DAC code: [LSB] 170
DAC A iref: [uA] 1.0
DAC B iref: [uA] 10.0
Plot: Frame number
Monitor every: 10
Refresh: 2
Current ev: 0
Time stamp 0

Figure 9.2: Overview of the core components of the DAQ software framework (a) and a screenshot of the GUI (b).

values and biasing conditions.

Once data taking begins, raw data from the DAQ board is passed to the EventBuilder, which translates the raw data in the form of 32-bit words into *CE-65 frames*. These are data structures consisting of metadata, including the frame number and timestamp, as well as the pixel signal in the form of the raw amplitude and CDS signal for every pixel in the chip matrix. The frames are subsequently packaged into a *CE-65 event*, consisting of 8 consecutive frames.

The EventBuilder then passes CE-65 events to the OnlineMonitor, where the data is used to create real-time visualisations for the user appearing in the GUI. These take the form of histograms embedded in the GUI's canvas. These are periodically updated throughout data taking.

The EventBuilder also passes CE-65 events to the DataStorage module. The core purpose of the DataStorage module is saving of CE-65 events in the form of .root files for offline analysis. The DataStorage module implements thresholding of CE-65 events for output data size reduction. The thresholds are user defined and required that a given CE-65 event contain a pixel whose signal (s_{ij}) exceeds

- $s_{ij} > 800$ ADUs if $s_{ij} \in$ AC submatrix
- $s_{ij} > 800$ ADUs if $s_{ij} \in$ DC submatrix
- $s_{ij} > 200$ ADUs if $s_{ij} \in$ SF submatrix

and that simultaneously the sum of the pixel and its adjacent pixels exceeds

- $S_{ij} = \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} s_{kl} > 3000$ (2000) ADUs if $s_{ij} \in$ AC submatrix of CE-65v1 (CE-65v2) chip
- $S_{ij} > 3000$ ADUs if $s_{ij} \in$ DC submatrix
- $S_{ij} > 1000$ ADUs if $s_{ij} \in$ SF submatrix

in order for the CE-65 event to be saved. The thresholds resulted in a data size reduction of ×12.

9.2 Noise Characterisation

The noise performance of the CE-65v1 and CE-65v2 chips was evaluated by measuring the chip responses in the absence of a radioactive source. Specifically, the pixel baseline and the equivalent noise charge (ENC) were analysed. The baseline represents the raw pixel response in the absence of a source.

Figure 9.3 depicts the pixel baseline for the 15 µm pitch CE-65v1 chip produced in the Modified process. The three submatrices corresponding to the AC-coupled preamplifier, DC-coupled preamplifier, and source-follower amplification schemes, introduced in Section 8.3, are clearly visible. The right-most SF submatrix displays a positive baseline in the $\sim$ 2500 ADU range, in accordance with simulations. The AC and DC submatrices display negative baselines, owing to the inverting amplifier.

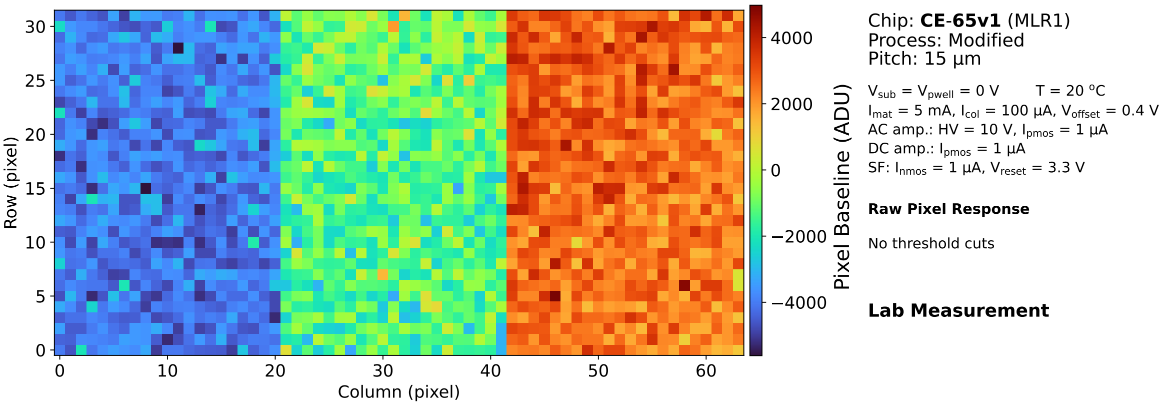
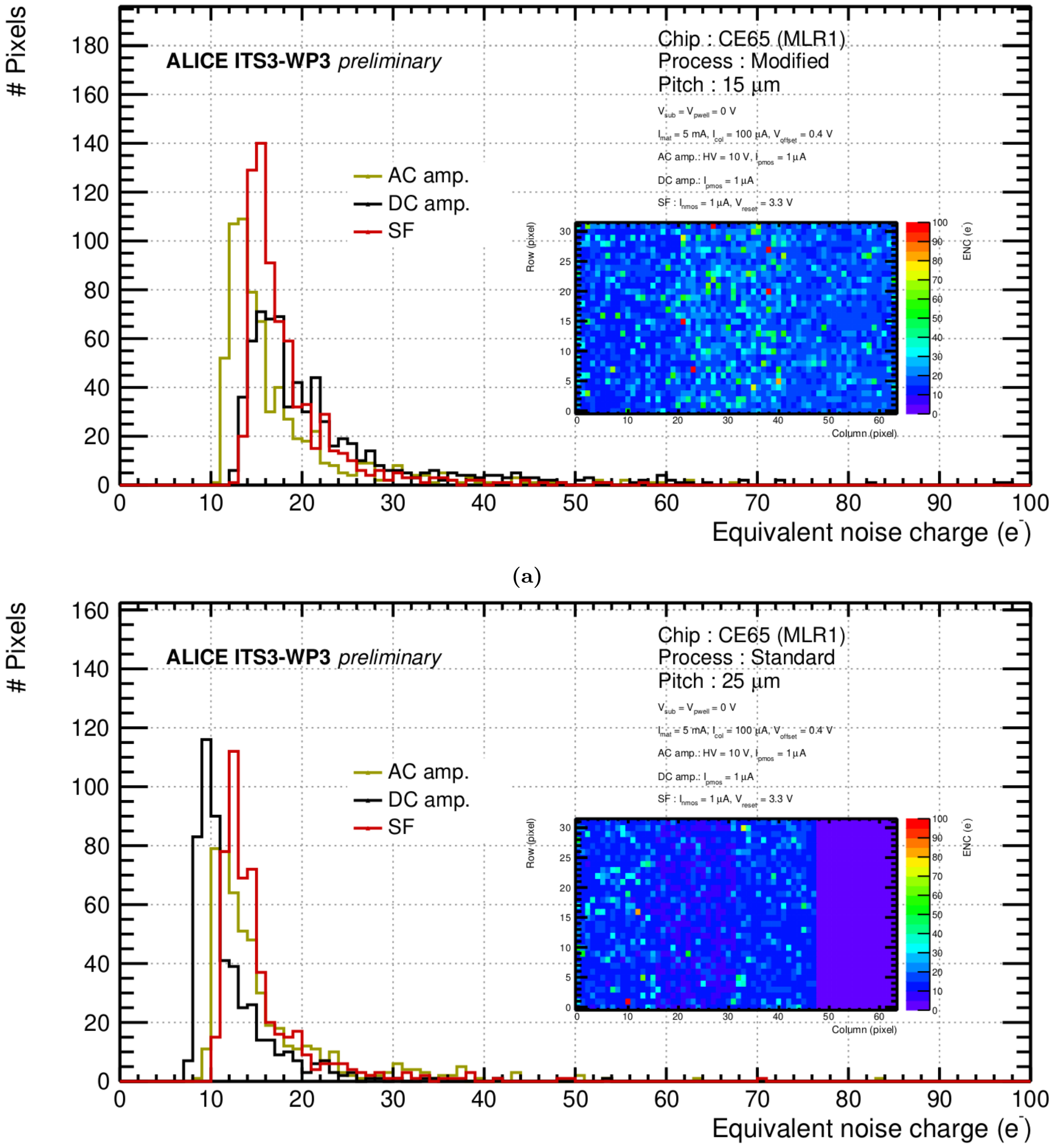


Figure 9.3: Baseline pixel response of the entire matrix of the 15 µm pitch CE-65v1 chip produced in the Modified process. Biasing information is displayed to on the right.

Equivalent Noise Charge Dependence on Chip Design

The ENC was computed by considering the CDS signal, introduced in Section 8.4, whereby the raw response of time-adjacent frames is subtracted. The noise distribution for each of the pixels was fit using a Gaussian. The standard deviation was then converted into electrons using the charge conversion factors computed in Section 9.6.3. Figure 9.4a shows the ENC distributions of the three submatrices of the 15 µm pitch CE-65v1 chip produced in the Modified process. The vast majority of pixels lie in the $\sim$ 10 to 20 electron range, with the AC submatrix displaying the lowest ENC, followed by the SF submatrix, as evidenced by the overlayed 2-d distribution of the ENC. All three distributions display the characteristic shape of a decaying exponential leading to a quickly decreasing number of noisy pixels. This is associated with rare fabrication defects or increased leakage currents, increasing noise in a small subset of pixels.

Figure 9.4b shows the ENC distributions of the three submatrices of the 25 µm pitch CE-65v1 chip produced in the Standard process for comparison. There is a clear decrease in the ENC of all three submatrices. The most drastic improvement is for the DC submatrix, which goes from being the noisiest submatrix in the Modified with Gap process chip, to the least noisy in the Standard process. A similar trend is observed when comparing the 15 µm pitch CE-65v1 chip produced in the Standard process to its Modified with Gap counterpart, suggesting that the increase in noise observed for the Modified and Modified with Gap chips is a property of the n-type blanket implant, that is magnified for the DC submatrix. The increase in noise can be associated with the increased capacitance resulting from the incomplete depletion of the low dose n-type implant around the collection electrode, as mentioned in Section 8.2. The effect is expected to be largest for the DC-submatrix, due to the low biasing voltage.

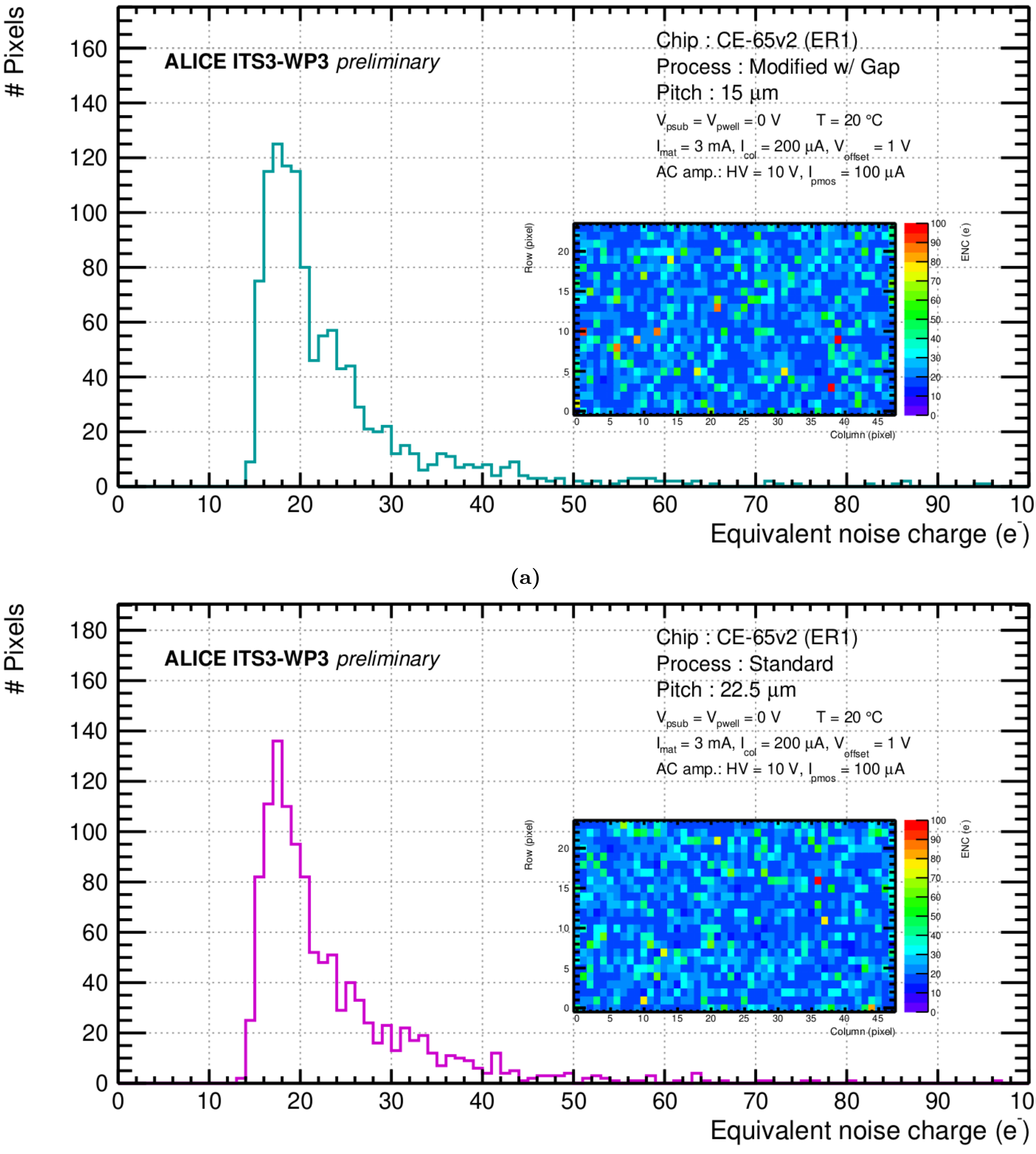


(a)

(b)

Figure 9.4: Equivalent Noise Charge distribution of the three submatrices of the CE-65v1 chip in the Modified process with a 15 µm pitch (a) and Standard process with a 25 µm pitch (b). The 2-d distribution is overlayed.

Figures 9.5a and 9.5b show the ENC distribution of the 15 µm pitch CE-65v2 chip produced in the Modified with Gap process and the 22.5 µm pitch CE-65v2 chip produced in the Standard process, respectively. The AC-coupled preamplifier is adopted for all pixels. A similar distribution to the ENC distributions of the CE-65v1 chips is evident, albeit with a higher ENC. The precise origin of this increase in noise is not known, but likely related to the different front-end. Comparing the distributions of the two CE-65v2 chips suggests little dependence on either pitch or process modification.



(a)

(b)

Figure 9.5: Equivalent Noise Charge distribution of the CE-65v2 chip in the Modified with Gap process with a 15 µm pitch (a) and Standard process with a 22.5 µm pitch (b). The 2-d distribution is overlayed.

9.3 Methodology of Radioactive Source Characterisation

9.3.1 Fe-55 Emission Lines

^{55}Fe is an isotope commonly used for the calibration of MAPS. The decay of ^{55}Fe proceeds overwhelmingly via electron capture, whereby an electron from one of the outer shells is captured by the nucleus:

$$^{55}_{26}\mathrm{Fe} + e^- \rightarrow ^{55}_{25}\mathrm{Mn} + \nu_e \tag{9.1}$$

The resulting Mn atom remains in an excited state with one of the inner shells missing an electron (typically the K-shell). This is compensated either through the emission of an electron, known as an Auger electron ($\sim$ 60%), or an electron transition from a higher shell, resulting in the emission of an X-ray ($\sim$ 30%)[211]:

$$^{55}_{25}\mathrm{Mn}^* \rightarrow ^{55}_{25}\mathrm{Mn} + \gamma \tag{9.2}$$

Table 9.1: The ^{55}Fe X-ray emissions and their Energies. Values adapted from Ref. [211].

Type	Transition	Energy (keV)	Relative Probability
$\mathbf{X_K}$	Kα_2	5.88772	51
	Kα_1	5.89881	100
	Kβ_1	6.49051	20.5
	Kβ_5''	6.5354	-included above-
$\mathbf{X_L}$	Ll	0.5576	–
	Lα	0.6394 – 0.6404	–
	Lη	0.5695	–
	Lβ	0.64636 – 0.7694	–
	Lγ	0.65826	–

The energy of the emitted X-rays is determined by the energy difference between the higher and lower shells. Common X-ray energies are summarized in Table 9.1, where it can be seen that the primary emission lines are determined by the transitions of electrons from the L-subshells to the K-shell, denoted as the K_{α_1} and K_{α_2} emissions. These occur with relative probability of 2:1 at energies of 5.89881 keV and 5.88772 keV, respectively [211]. The energy difference is far smaller than the intrinsic energy resolution of silicon detectors [212]:

$$R = 2.35 \cdot \sqrt{\frac{Fw}{E}} \approx 2\% \tag{9.3}$$

where F is the Fano factor in silicon, w is the excitation energy of an electron-hole pair in silicon, and E is the energy deposited by an impinging X-ray. In effect the K_{α_1} and K_{α_2} emissions result in a single peak, denoted as the "K_α peak" throughout this thesis. In addition, Table 9.1 highlights also transitions from the M-subshells to the K-shell, denoted by Kβ_1 and Kβ_5''. The energy difference is once again smaller than the intrinsic energy resolution, resulting in a single K_β peak with $\sim$ 8 times less counts than its K_α counterpart.

9.3.2 Clusterisation

The charge deposited by an impinging X-ray of the decaying ^{55}Fe source is seldom collected by a single pixel, but rather diffuses to neighbouring pixels where it is collected. This is particularly the case for Standard process chips, where the lateral parts of the epitaxial region remain undepleted resulting in the absence of an electric field to cause electrons to drift to the collection diode. Thus it becomes necessary to consider not only the central pixel upon which the X-ray impinges, the *seed*, but also its adjacent pixels, the *neighbours*, during the construction of the ^{55}Fe spectrum. This reconstruction procedure is dubbed *clusterisation*, and is explored below.

Table 9.2: Clusterisation thresholds for different chip variants in ADUs.

Chip Type	Amp. Scheme	Seed Threshold	Neighbour Threshold
CE-65v2	AC-PreAmp.	1000	300
CE-65v1	AC-PreAmp.	1000	300
	DC-PreAmp.	800	200
	Source-Follower	200	60

Seed Pixel Spectrum

Formally the seed pixel is defined as any pixel that passes a given ADU seed threshold. These are summarized in Table 9.2 for the different chip configurations that were tested. The seed threshold corresponds to ~ 9 times the mean RMS noise of each pixel in the three CE-65v1 submatrices. The threshold is implemented in order to reject the large quantity of noise hits, and hits which may originate from the charge of an X-ray impinging on one of the central pixel's neighbours and diffusing across pixel boundaries, rather than the pixel itself. A typical seed pixel spectrum for the 15 µm CE-65v2 chip in the Modified with Gap process is presented in Figure 9.6a. The main K_α peak is clearly visible at ~ 6900 ADUs corresponding to the K_α emission line, often referred to as the calibration peak.

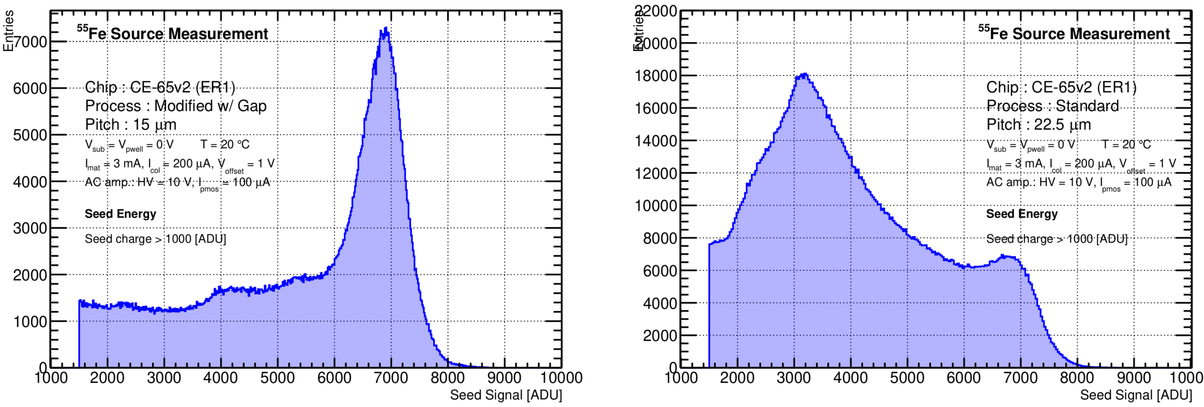


Figure 9.6: The ^{55}Fe spectrum for the seed signal of the entire matrix of the CE-65v2 chip in the Modified with Gap process with a 15 µm pitch (a) and Standard process with a 22.5 µm pitch (b).

Figure 9.6b displays the seed pixel spectrum for the 22.5 µm CE-65v2 chip in the Standard process. Although the main K_α peak is once again visible at ~ 6800 ADUs, the spectrum is dominated by another peak at ~ 3200 ADUs, known as the charge sharing peak. The charge sharing peak corresponds to events where the impinging X-ray deposits less than half of the corresponding charge in the seed pixel and the rest of the charge diffuses to neighbouring pixels.

Single-Pixel Cluster

The single-pixel cluster spectrum is defined as the subset of events where the seed pixel passed the ADU seed threshold listed in Table 9.2, but none of the adjacent neighbours pass the ADU neighbour threshold. The neighbour threshold corresponds to approximately 2 times the mean RMS noise of each pixel, and is implemented in order to reject positive noise fluctuations. When choosing the neighbour threshold a trade-off is made between noise rejection and rejection of legitimate signal resulting from charge sharing.

The core principle of the single-pixel clusteristation procedure is to isolate events where the overwhelming majority of the charge of impinging X-rays is collected by the seed pixel, in order to calibrate the ADU response of the sensor to the K_α peak. This is especially important for chips with high charge sharing, where the K_α peak may be less prominent or not visible at all in the seed spectrum. Figure 9.7a depicts the single-pixel spectrum for the 22.5 µm Standard process CE-65v2 chip. The K_α peak can easily be distinguished and lower ADU contributions of seed pixels due to charge sharing have been sizeably reduced.

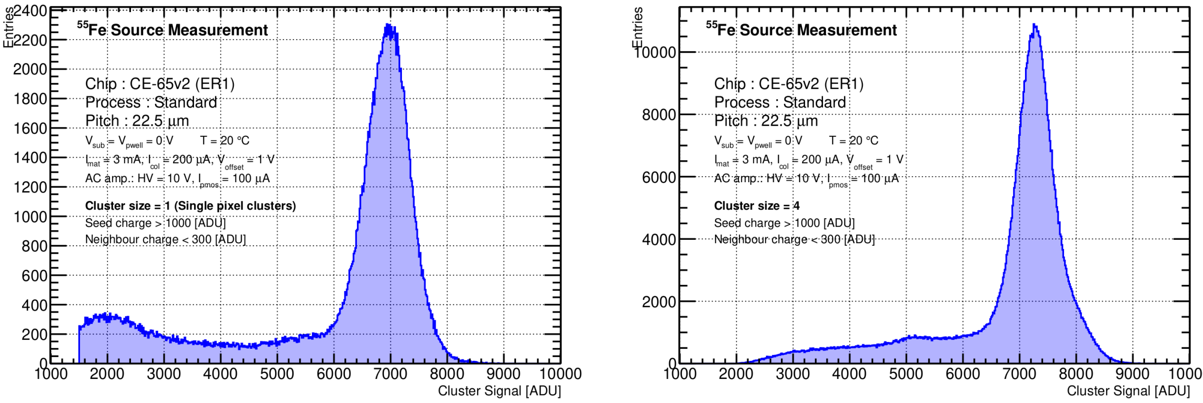


Figure 9.7: The ^{55}Fe spectrum for the single-pixel clusters (a) and size-4 clusters (b) of the entire matrix of the CE-65v2 chip in the Standard process with a 22.5 µm pitch (b).

Size-n Cluster

The clusterisation procedure introduced above can be defined to iteratively include not only adjacent neighbours, but also neighbours of the neighbours, and so on. After a seed pixel is identified by passing the ADU seed threshold, all adjacent pixels to the seed pixel are required to pass the ADU neighbour threshold. For each neighbour passing the neighbour threshold, the pixel is added to the cluster and the procedure is repeated, such that each of the pixels adjacent to the neighbours having been defined in the previous step, are once again required to pass the neighbour threshold. This is repeated iteratively until there are no pixels adjacent to the defined cluster that pass the neighbour threshold. The seed of the resulting cluster is defined as the most energetic pixel, rather than the original pixel passing the seed threshold where the clusterisation procedure started, although in practice these are often the same. The cluster is then defined as a size-n cluster if there n pixels in the cluster. For the purposes of this thesis, only cluster sizes < 8 were considered.

Figure 9.7b presents the size-4 cluster for the 22.5 µm Standard process CE-65v2 chip. The spectrum is relatively devoid of low-ADU noise and charge sharing contributions, being dominated by a prominent K$_\alpha$ peak. The most interesting application of the size-n spectra is the comparison among different cluster sizes as a gauge for the charge sharing of the different chip configurations. This is explored in Section 9.5.3.

Matrix Signal

The matrix signal is the last clusterisation strategy introduced in this thesis and simply describes the addition of the signal in a seed pixel and all its adjacent pixels in a window. It was introduced as S_{ij} in Section 9.1.1 as one of the online thresholding criteria. Throughout this thesis, window sizes of 3×3 and 5×5 were considered. Figure 9.8 depicts the 3×3 matrix signal for the 15 µm Modified with Gap process and 22.5 µm Standard process CE-65v2 chips.

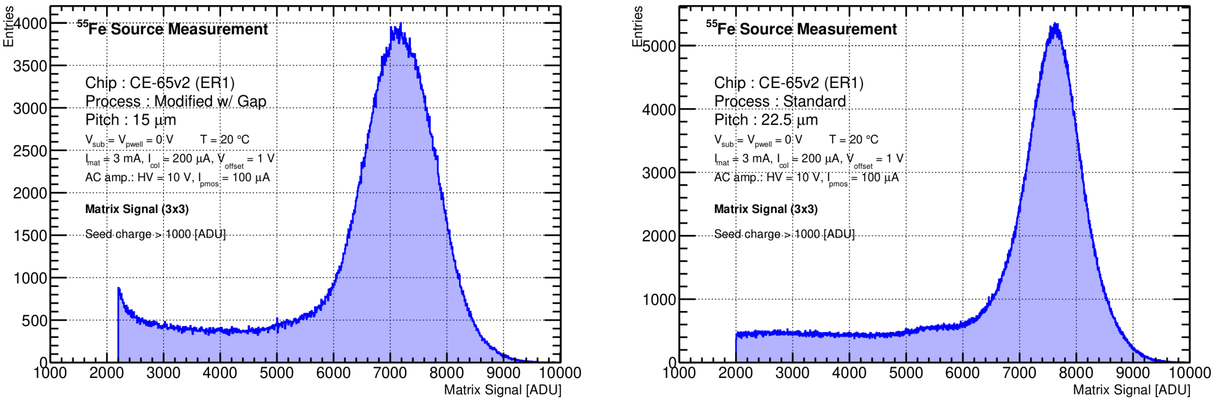


Figure 9.8: The ^{55}Fe spectrum for the matrix signal (3×3) of the entire matrix of the CE-65v2 chip in the Modified with Gap process with a 15 µm pitch (a) and Standard process with a 22.5 µm pitch (b).

While the spectrum is expected to contain the complete charge deposited by an impinging X-ray, it also introduces the noise of the $n^2 - 1$ pixels surrounding the seed, resulting in a smeared peak that can be seen at $\sim$ 7200 ADUs and $\sim$ 7600 ADUs, respectively. Given that the entire charge of the impinging X-rays from the ^{55}Fe source should be contained in the matrix signal, the long tail to the left of the Gaussian can be identified as originating not only from lower energy emission lines (e.g. Si-K$_\alpha$), but also from conversions in the p$^+$ substrate, where the generated charge is not fully collected [212]. The matrix signal was used as a proxy for estimating the complete collected signal, and comparing it to the collected signal for individual pixels. A detailed comparison is presented in Section 9.5.2 for the different sensor variants.

9.3.3 Fitting Methodology

To model the K$_\alpha$ peak appearing in the different ^{55}Fe spectra a Gaussian fit was used. In spectra where the K$_\beta$ peak was discernible, such as in Figure 9.17a, a sum of two Gaussians was used:

$$I(x_{\mathrm{ADU}}) = A \cdot \exp\left(-\frac{(x_{\mathrm{ADU}} - \mu_\alpha)^2}{2\sigma_\alpha^2}\right) + B \cdot \exp\left(-\frac{(x_{\mathrm{ADU}} - \mu_\beta)^2}{2\sigma_\beta^2}\right) \tag{9.4}$$

where x_{ADU} represents the measured voltage in ADUs, μ_i is the Gaussian mean, and σ_i^2 is the variance. As the K$_\alpha$ peaks dominate the ^{55}Fe spectra, no terms were needed to describe the background. The fit range was chosen around the most-probable value (MPV) of the distribution as MPV $\pm 0.05 \cdot$ MPV, but was adapted where necessary if the fit did not properly describe the underlying distribution.

A set of constraints was imposed on the fit parameters in order to ensure convergence (and where appropriate fit the K$_\beta$ peak). The constraints for the AC submatrix are reported below. Other submatrices differed only in the peak width owning to their differing energy resolutions.

- **K$_\alpha$ Peak Position ($\boldsymbol{\mu_\alpha}$):** The K$_\alpha$ peak position was required to lie within MPV $\pm$ MPV/12. In practice this constraint was purely for fit stability as the MPV lies on the K$_\alpha$ peak.
- **K$_\beta$ Peak Position ($\boldsymbol{\mu_\beta}$):** The K$_\beta$ peak position was required to lie within $[1.05 \cdot \mu_\alpha, 1.15 \cdot \mu_\alpha]$. This constraint was imposed to center the μ_β around the expected nominal K$_\beta$ peak position of $\sim 1.10 \cdot \mu_\alpha$.
- **Peak Widths ($\boldsymbol{\sigma_i}$):** The K$_\alpha$ and K$_\beta$ peak width was required to lie within $[100 \text{ ADUs}, 10000 \text{ ADUs}]$. As 10000 ADUs spans the entire spectrum this was in effect only a lower bound of $>\sim 1\%$.
- **K$_\beta$ Amplitude ($\boldsymbol{B}$):** The K$_\beta$ amplitude was required to lie within $[0, 0.2 \cdot \text{MPV}]$ reflecting the relative transition probability given in Section 9.3.1 ($A/B \sim 8$).

In spectra where the K$_\beta$ peak could not be discerned, only the K$_\alpha$ contraints given above were applied. The fitting procedure was performed using the least-squares method, with the χ^2 serving as the goodness-of-fit estimate.

Charge Conversion Factor

The most prominent K$_\alpha$ peak, introduced in Section 9.3.2, was used to perform a mapping from the measured ADUs to energy by matching the known peak positions of the K$_\alpha$ X-rays. The energy is often expressed in terms of electrons, corresponding to the number of electron-hole pairs produced by the impinging X-ray. In silicon the mean energy for the creation of an electron-hole pair is 3.6 eV, meaning that an X-ray of 5.9 keV corresponding to the K$_\alpha$ emission line will on average generate 1640 electron-hole pairs. The mapping is defined as the charge conversion factor (γ_{ccf})

$$\gamma_{\text{ccf}} = \frac{dQ}{dE} = \frac{N_e \cdot e}{\mu_\alpha \cdot \text{ ADU}} = \frac{1640}{\mu_\alpha} \cdot \frac{e}{\text{ADU}} \tag{9.5}$$

where Q is the generated charge in the sensing volume, E is the energy of the impinging X-ray, and e is the electron charge. This computation assumes a complete charge collection, which was found to hold reasonably for most sensor variants, as detailed in Section 9.5.2.

9.4 Gain Calibration

An extensive characterisation of the CE-65v1 and CE-65v2 chips was performed using X-rays from an ^{55}Fe source. Each of the different flavours of the CE-65v2 chip were measured at constant 20^{o}C using a chiller for temperature control. CE-65v1 chips were measured at room temperature (i.e. without cooling). The single-pixel cluster spectrum, introduced in Section 9.3.2, was used for calibration under the assumption that the vast majority of the charge was collected by a single pixel. As shown below, this is the case for the Modified and Modified with

Gap process chips, but breaks down for the DC-coupled preamplifier of the Standard process CE-65v1 chip. Thus for the DC submatrix of the Standard process CE-65v1 chip the matrix signal was used instead, albeit at the cost of smearing the individual pixel calibrations due to the increased noise.

Dependence on Process Modification and Pixel Pitch

Figure 9.9a depicts the measured ^{55}Fe spectrum for single-pixel clusters for the CE-65v2 chip in the Modified with Gap process with a 15 µm pitch. The main K_α peak is centered around $\sim$ 6900 ADUs. The neighbouring K_β peak, which would be expected at $\sim$ 7480 ADUs, cannot be resolved due to pixel-to-pixel gain variations detailed in Section 9.6. The procedure was repeated for the 22.5 µm pitch Modified with Gap CE-65v2 chip, as well as the 15 µm and 22.5 µm chips in the Standard process, yielding similar spectra. The K_α peaks were centered around $\sim$ 7040 ADUs, $\sim$ 6800 ADUs, and $\sim$ 6980 ADUs, respectively. Figure 9.9b depicts the ^{55}Fe spectrum for single-pixel clusters for the CE-65v2 chip in the Standard process with a 22.5 µm pitch for comparison. The peaks lie within 3.4% of each other, suggesting only a weak dependence on both process modification and pixel pitch.

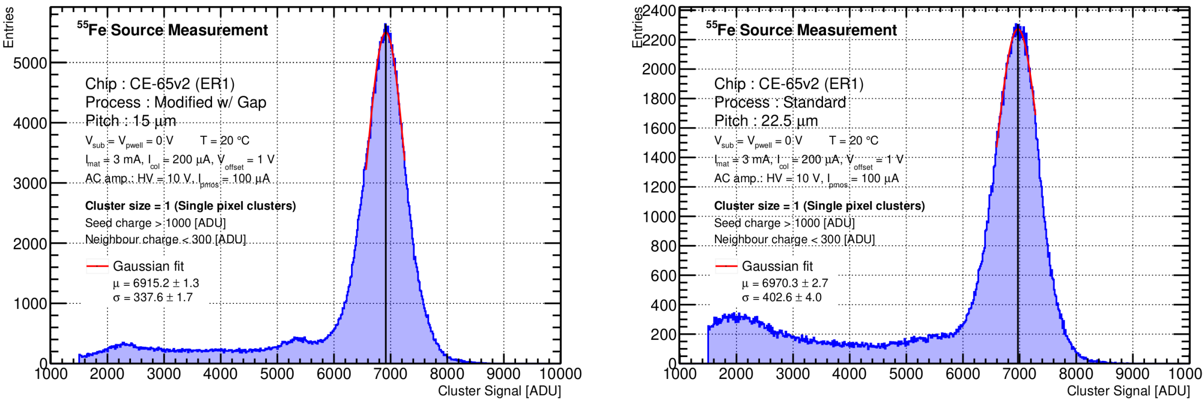


Figure 9.9: The ^{55}Fe spectrum for single-pixel clusters of the entire matrix of the CE-65v2 chip in the Modified with Gap process with a 15 µm pitch (a) and Standard process with a 22.5 µm pitch (b). The Gaussian fit is marked in red.

This outcome is anticipated, as the capacitance should not vary significantly when the chips are fully depleted. Nevertheless, it should be noted that a slight increase in peak position is observed both when moving from the 15 µm to the 22.5 µm pitch, and from the Standard Process to the Modified with Gap process.

Dependence on Amplification Scheme

Significant differences are observed in the K_α peak positions across the three submatrices of the 15 µm pitch CE-65v1 chip in the Modified process. The AC submatrix registers the highest peak at 7120 ADUs, while the DC submatrix follows with a slightly lower value of 6120 ADUs. The SF submatrix, depicted on Figure 9.10a, has the lowest peak at 1910 ADUs, owing to its source-follower architecture, which does not amplify the signal.

The 25 µm pitch CE-65v1 chip in the Standard process shows largely similar trends, with the AC submatrix K_α peak position again at a similar position at $\mu \approx 7290$ ADUs. The biggest difference can be seen for the DC submatrix, depicted on Figure 9.10b, where the peak position is now the highest at $\mu \approx 7530$ ADUs. The increase with respect to its AC counterpart is explained by the parasitic capacitance introduced by the AC coupling. The SF submatrix again displays the smallest peak position at $\mu \approx 2360$ ADUs, although considerably higher than in the Modified process chip. The increase in the peak position for both the SF and DC submatrices with respect to their counterparts in the 15 µm pitch Modified process chip can be explained by the faster depletion of the Standard process, leading to a lower capacitance. The effect is not observed for the AC submatrix, since in for both chips depletion voltage of 10V ensures the chips are fully depleted. This is explored in detail in Section 9.4.

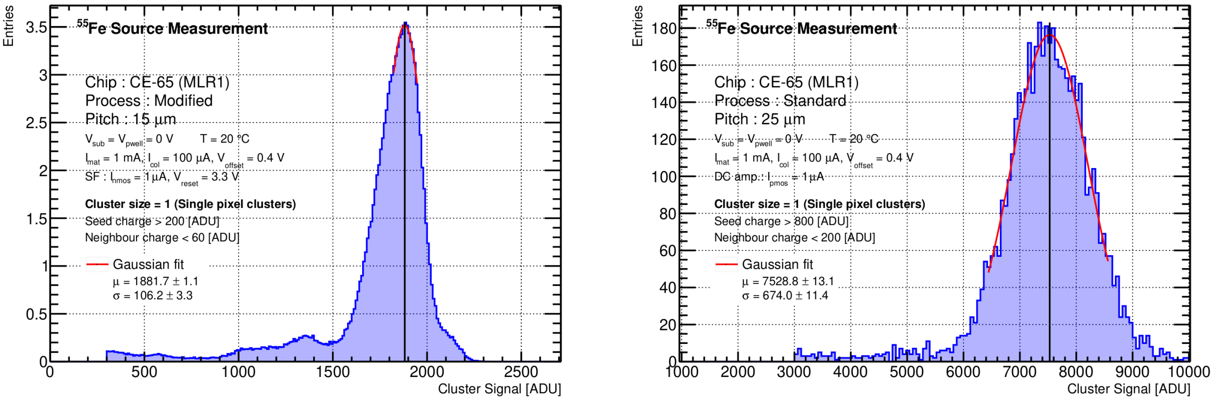


Figure 9.10: The ^{55}Fe spectrum for single-pixel clusters of the SF submatrix of the CE-65v1 chip in the Modified process with a 15 µm pitch (a) and the DC submatrix in the Standard process with a 25 µm pitch (b). The Gaussian fit is marked in red.

Dependence on Depletion

The depletion of the sensor volume was studied by applying different biasing voltages to the collection node, as introduced in Section 8.3. Figure 9.11 depicts the evolution of the charge conversion factor (γ_{ccf}) as a function of biasing voltage. As the biasing voltage is increased the capacitance quickly decreases resulting in a larger voltage response to the impinging K_α X-rays. In turn, this results in a quickly decreasing charge conversion factor until a biasing voltage of $\sim$ 4V, after which saturation is reached and the depletion region stops developing. This corresponds to the voltage at which the depletion region develops vertically until the p+ substrate is reached, and the high doping gradient prevents further depletion.

The chips produced in the Standard process achieve depletion considerably more quickly, evidenced by the relatively low charge conversion factor already achieved with no reverse biasing voltage applied. This contrasts heavily with the Modified with Gap process chips which require a biasing voltage of at least 2V to achieve a similar capacitance. This can be attributed to the incomplete depletion of the low dose n-type implant, requiring a higher biasing voltage to achieve a similar depletion around the collection electrode. At a biasing voltage of 4V and above, the four CE-65v2 chips display a similar charge conversion factor, as expected.

The equivalent result for the CE-65v1 chips, depicted on Figure 9.12, tells a similar story. The AC submatrix of the 25 µm pitch CE-65v1 chip in the Standard process depletes very similarly to the Standard process CE-65v2 chips, with the chips achieving virtually full depletion already at a depletion voltage of 2V. The AC submatrix of the 15 µm pitch CE-65v1 in the Modified process depletes slightly more slowly than its its Modified with Gap process CE-65v2 counterpart. In particular, the chip does not achieve full depletion until about 10 V.

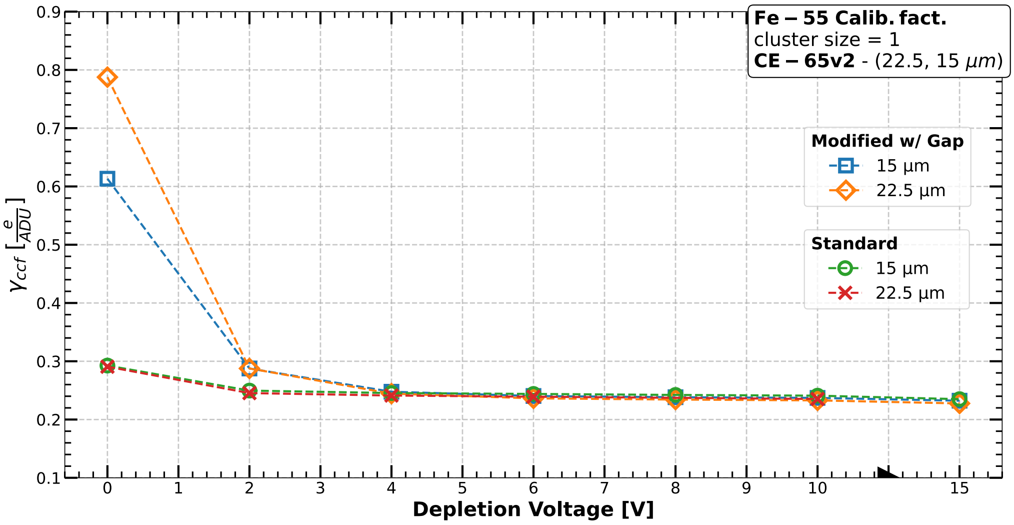


Figure 9.11: Charge conversion factor (γ_{ccf}) as a function of biasing voltage for the four tested variants of the CE-65v2 chip.

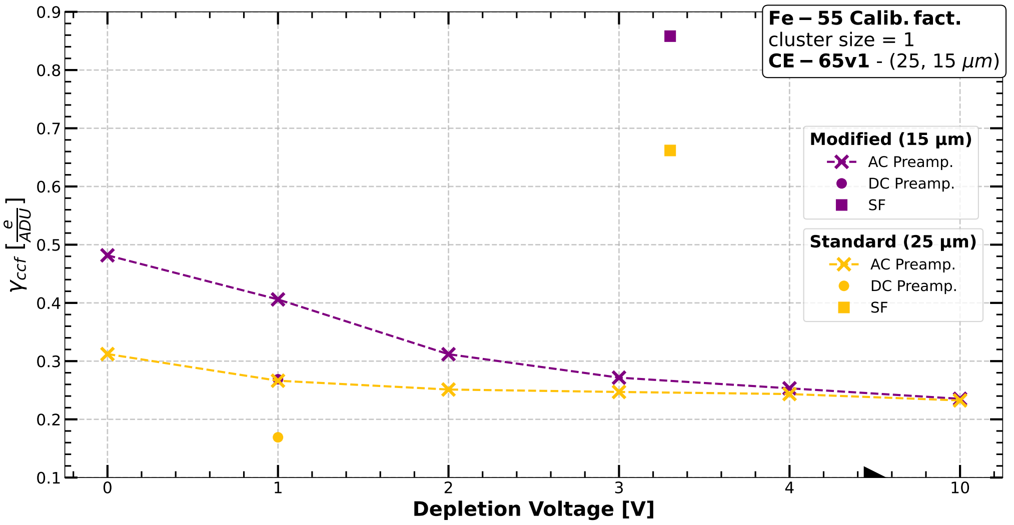


Figure 9.12: Charge conversion factor (γ_{ccf}) as a function of biasing voltage for the two tested variants of the CE-65v1 chip.

9.5 Charge Sharing

Charge sharing positively impacts the achievable spatial resolution of sensors, but also makes sensor less robust with respect to radiation defects and noise. It was found to depend strongly on process modification and amplification scheme. This section explores the charge sharing of different CE-65v1 variants through the fractional charge carried by neighbours, the charge collection efficiency, and the cluster size composition of the ^{55}Fe spectrum.

9.5.1 Seed vs Neighbours

The distribution of the energy carried by the seed and by the neighbours in a given event passing the seed thresholds listed in Table 9.2 is shown in Figures 9.13a and 9.13b for the AC submatrix of the 15 µm Modified process and the DC submatrix of the 25 µm Standard process CE-65v1 chips, respectively. Each histogram visualizes the charge distribution within clusters resulting from ^{55}Fe X-rays, where the x-axis corresponds to the charge carried by the seed pixel, and the y-axis to the combined charge of neighbouring pixels. The total cluster charge in a given event is thus given by $x + y$.

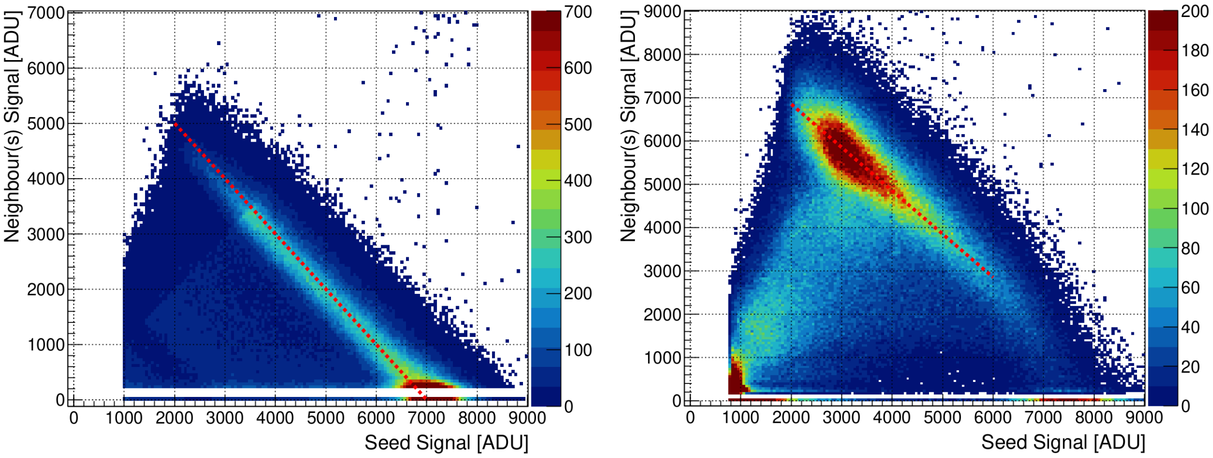


Figure 9.13: Distribution of seed energy vs. neighbour energy in the ^{55}Fe spectrum of the AC submatrix of the CE-65v1 chip in the Modified process with a 15 µm pitch (a) and the DC submatrix in the Standard process with a 25 µm pitch (b). Red lines are overlaid at $x+y = 7000$ and $x + y \approx 8850$ for visual reference. The colour bar is capped; some bins exceed the displayed maximum.

Figure 9.13a shows a clear diagonal along the $x + y = 7000$ line corresponding to signal hits resulting from the K_α peak. The vast majority of signal hits lie in the low charge sharing regime, with the cluster seed typically carrying more than 6000 ADUs. The $x = y$ line that is faintly visible corresponds to the "decision boundary" of 2-pixel events, where the most energetic pixel is considered the seed, leading to an abruptly higher density of events below the $x = y$ line.

By comparison, the distribution of the DC submatrix of the 25 µm Standard process CE-65v1 chip in Figure 9.13b exhibits a considerably higher signal region, with most events lying along the $x + y =\sim 8850$ line. The vast majority of events now lies in the high charge sharing regime, with the seed pixel carrying a small fraction of the signal. Typically this is less than 4000 ADUs. Interestingly, the low charge sharing events do not lie on the $x + y =\sim 8850$ line, but tend to lower ADU counts. The events with little or no cluster energy carried by neighbours peaks at $\sim$ 7500 ADUs, which is reflected in Figure 9.10b.

The large discrepancy in the peak position between the high and low charge sharing events in the case of the DC submatrix of the 25 µm Standard process CE-65v1 chip meant that the DC submatrix was exceptionally calibrated using the 5×5 matrix signal distribution in order to include the full contribution of neighbouring pixels. In effect, this means that the nominal calibration values presented in Section 9.4 are likely significant underestimates. Nevertheless, they still allow for the pixel-to-pixel gain variations to be studied.

For all other measured CE-65v1 chip configurations the discrepancy between the peak positions was considerably smaller ($< 5\%$).

9.5.2 Charge Collection Efficiency

The charge collection efficiency is introduced in Ref. [212] as the ratio between the K_α peak position determined from the 3×3 matrix signal ^{55}Fe distribution, and the peak position determined from the single-pixel spectrum. Under the assumption that the single-pixel spectrum is dominated by events where the charge conversion from impinging X-rays takes place entirely within the depletion region of the sensing volume, it is expected that the measured ADU count corresponds to the full 1640 electrons generated. The assumption was found to hold best for the AC submatrix of the CE-65v1 chips, as the large biasing voltage allows for a full depletion of the sensing volume (vertically).

The matrix signal, on the other hand, contains contributions from events where the charge is split between more than 1 pixel, either because the charge conversion took place in an undepleted part of the epitaxial layer, or because a part of the charge conversion took place in the p^+ substrate. In either case, the charge will not drift directly to the collection electrode of the pixel in which it was generated. In the latter case, the electrons may not diffuse into the epitaxial layer, leading to an incomplete charge collection. This results in a shifting of the K_α peak position towards lower ADU counts. This effect is reproduced in Refs. [213] and [212]. Thus, the charge collection efficiency gives an estimate of the fraction of generated charge that is actually collected.

Figure 9.14 depicts the charge collection efficiency as a function of biasing voltage for the AC submatrices of the 15 µm Modified process and 25 µm Standard process CE-65v1 chips. In order to disentangle the influence of charge sharing between the epitaxial layers of neighbouring pixels, the charge collection efficiency for both the 3×3 and 5×5 matrix signal was included.

The charge collection efficiency increases monotonously for the Standard process chip as higher biasing voltages are applied and the depletion region develops. Comparing the 3×3 to the 5×5 charge collection efficiency reveals that the majority of this effect is due to charge diffusing beyond the immediate neighbours of the seed pixel. Nevertheless, even the 5×5 charge collection efficiency is found to gradually increase with the depletion voltage. At a depletion voltage of 10V a charge collection efficiency of $\sim 96\%$ is achieved.

The charge collection efficiency of the Modified process chip is above 100% at 0V, suggesting that the assumption that the single-pixel spectrum is dominated by events where the full 1640 electrons are collected does not hold at low biasing voltages. This is in agreement with the considerably slower depletion of the Modified process chip discussed in Section 9.4. At biasing voltages of 2V and above, the charge collection efficiency is relatively stable at above 97%. The charge collection efficiency shows virtually no dependence on the size of the matrix window, suggesting that the effect is primarily due to incomplete charge conversion in the epitaxial layer.

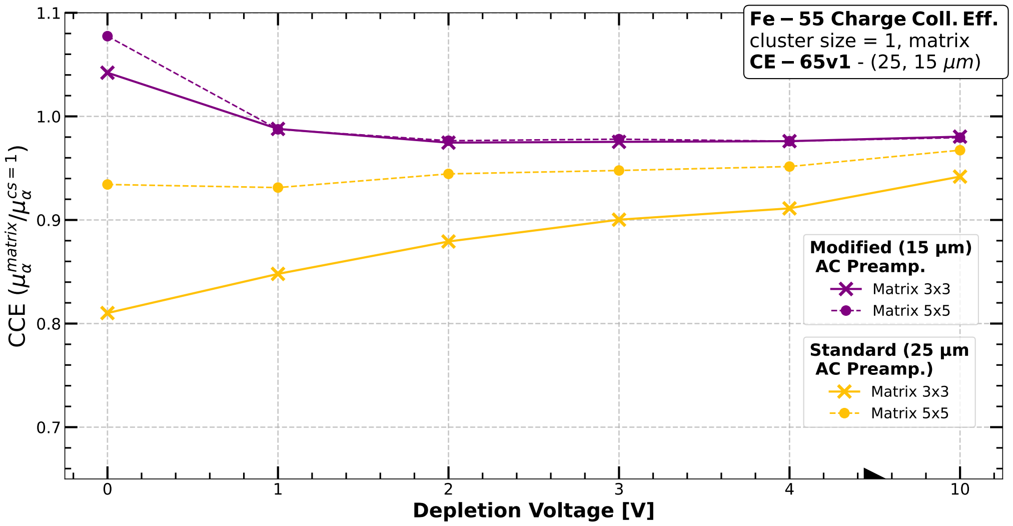


Figure 9.14: Charge collection efficiency ($\mu_\alpha^{matrix}/\mu_\alpha^{cs=1}$) as a function of biasing voltage for the two tested variants of the CE-65v1 chip.

9.5.3 Contributions of Varying Cluster Sizes

The clusterisation procedure described in Section 9.3.2 results in a variety of clusters corresponding to different cluster sizes. For the majority of this discussion the single-pixel cluster spectrum was considered, as this offers the most obvious proxy for the characterisation of the pixel response to the ^{55}Fe source. Nevertheless, considering the spectra of varying cluster sizes can be instructive as it offers insight into the charge collection of a given chip for a typical ^{55}Fe X-ray.

Figure 9.15a depicts the n-pixel cluster spectra of the 15 µm pitch CE-65v1 chip produced in the Modified process for cluster sizes $n = \{1, 2, 3, 4, 5, 6, 7\}$, as well as the overlap of all cluster sizes. The ^{55}Fe distribution is dominated by contributions from low cluster sizes, with single-pixel clusters being the largest contribution, and the contribution with respect to increasing cluster sizes quickly decreasing. As the cluster sizes are increased, a slight shift downward towards the measured matrix signal at $\sim$ 6970 occurs. For a cluster size of 4 the signal is found to lie below the matrix signal, suggesting that the noise cuts imposed in the clusterisation procedure exclude a small fraction of the signal.

Figure 9.15b depicts the n-pixel cluster spectra of the 25 µm pitch CE-65v1 chip produced in the Standard process. By comparison, cluster sizes below 4 do not contribute significantly to the ^{55}Fe spectrum, which is dominated by size 4 clusters, and size 5 clusters to a lesser degree. The K_α peak position varies greatly for the different cluster sizes. The single-pixel clusters have the highest K_α peak position at $\sim$ 7300 ADUs, which quickly falls to $\sim$ 6800 for cluster sizes 2 and 3. Cluster sizes of 3 and below, however, contribute very little to overall spectrum. There is a large drop in the K_α peak position for cluster sizes of 4 to $\sim$ 6400 ADUs, which gradually begins to increase as the cluster size is increased. This is well below the measured matrix signal K_α peak position of $\sim$ 7050 ADUs, suggesting that a large fraction of the signal is being excluded by the clusterisation procedure. Though this was also observed for the Modified variant, the effect is considerably more pronounced in the Standard process chip due to increased charge sharing. The distributions for cluster sizes of 4 and above no longer follow a pure Gaussian, but rather include an exponential tail towards higher ADU events.

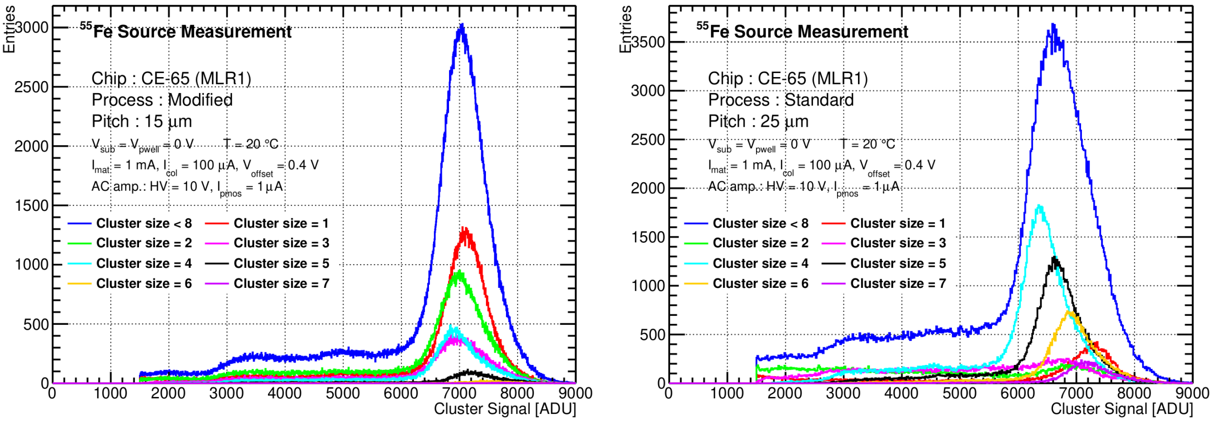


Figure 9.15: The ^{55}Fe spectrum for size-n clusters of the AC submatrix of the CE-65v1 chip in the Modified process with a 15 µm pitch (a) and Standard process with a 25 µm pitch (b). The blue distributions corresponds to the summation of all other distributions.

9.6 Gain Uniformity

The gain uniformity of CMOS imaging sensors can vary widely, and is one of the main disadvantages with respect to CCD sensors [214]. The large matrix size of the CE-65 family of chips enables a detailed, evaluation of the uniformity of the sensor response. This section presents gain calibration and energy resolution studies, highlighting both variations across the matrix and differences between chip variants.

9.6.1 Individual Pixel Spectra

The fitting procedure introduced in Section 9.3.3 was repeated for the single-pixel clusters of individual pixels, yielding much cleaner spectra, at the cost of limited statistics. Figure 9.16a depicts the single-pixel spectrum for an individual pixel (x=5, y=2) of the 15 µm pitch CE-65v2 chip produced in the Modified with Gap process. The energy resolution of 4.8%, obtained as the ratio of the Full Width at Half Maximum and mean peak position, is considerably better than that of the global spectrum in Figure 9.9a, where it is 11.5%. The K$_\beta$ peak can be clearly resolved, which would be expected to appear at $\sim$ 7480 ADUs, assuming linearity. Figure 9.16b depicts the single-pixel spectrum for an individual pixel (x=19, y=20) of the 22.5 µm pitch CE-65v2 chip produced in the Standard process. While the energy resolution remains low in accordance with Figure 9.9b, the high charge sharing means the statistics are extremely limited due to the rare occurrence of single-pixel clusters.

The procedure was also repeated for the single-pixel clusters of the individual pixels of the CE-65v1 submatrices. All three submatrices of the 15 µm CE-65v1 chip in the Modified process displayed narrow spectra with a good energy resolution. Figure 9.17a depicts the single-pixel spectrum for an individual pixel (x=59, y=28) of the SF submatrix, where an energy resolution of 5.7% can be observed. The K$_\beta$ peak was able to be fit only for a fraction of the SF submatrix pixels, as often the energy resolution was too poor.

Due to the high charge sharing of the SF and DC submatrices of the Standard process variant, a dedicated high-statistics run was conducted consisting of a week of data taking using the ^{55}Fe source introduced in Section 9.1.1. A shift in the single-pixel spectrum peak of approximately 3-4% to lower ADU counts was observed over long runs, attributable to gradual heating of the chip, which increased the noise level. Figure 9.17b depicts the single-pixel spectrum for an individual pixel (x=20, y=20) of the DC submatrix of the 25 µm pitch CE-65v1 chip in

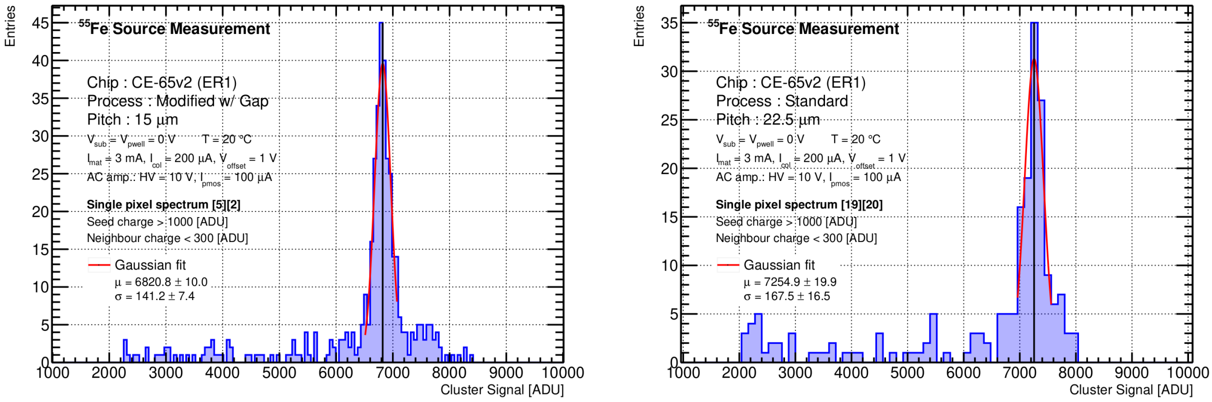


Figure 9.16: The ^{55}Fe spectrum for single-pixel clusters for an individual pixel ($x = 5$, $y = 2$) of the CE-65v2 chip in the Modified with Gap process with a 15 µm pitch (a). ^{55}Fe spectrum for single-pixel clusters for an individual pixel ($x = 19$, $y = 20$) of the CE-65v2 chip in the Standard process with a 22.5 µm pitch (b). The Gaussian fit is marked in red.

the Standard process. The spectrum not only displays very low statistics, but also an energy resolution of 22.1%, which is much worse than its Modified process counterpart.

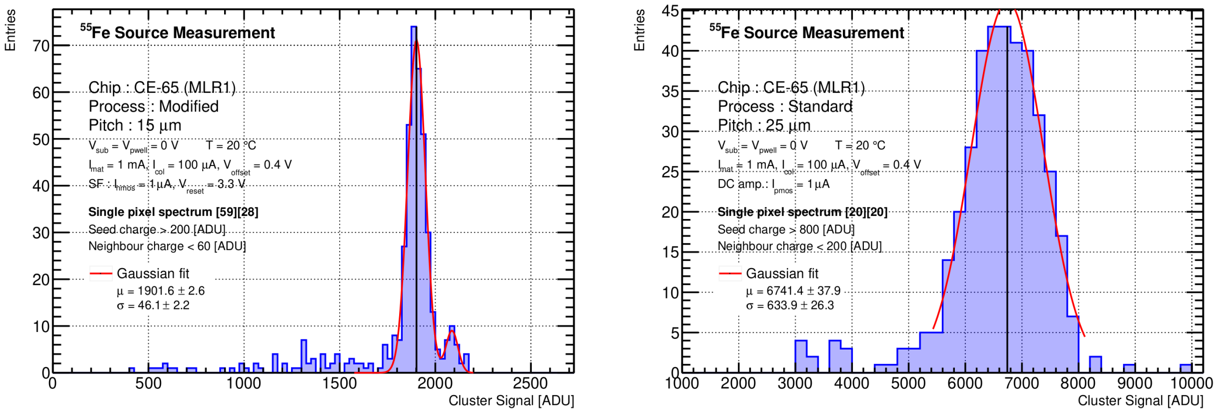


Figure 9.17: The ^{55}Fe spectrum for single-pixel clusters for an individual pixel ($x = 59$, $y = 28$) of the SF submatrix of the CE-65v1 chip in the Modified process with a 15 µm pitch (a) and for an individual pixel ($x = 20$, $y = 20$) of the DC submatrix in the Standard process with a 25 µm pitch (b). The Gaussian fit is marked in red.

9.6.2 Energy Resolution

A quantitative study of the energy resolution of each CE-65 chip variant was performed by considering the distribution of the energy resolution obtained from the Gaussian fit of each individual pixel.

Dependence on Process Modification and Pixel Pitch

Figure 9.18a depicts the normalized distribution of the energy resolution of each pixel in the 15 µm pitch CE-65v2 chip produced in the Modified with Gap process. The mean of the distribution corresponds to an energy resolution of 5.85%, with a relatively low standard deviation of 1.46%, suggesting that the vast majority of pixels achieve a good energy resolution. To the right of the distribution there is a distinct tail resembling the behaviour of a decaying exponential related to a quickly decreasing number of high-noise pixels that are smearing the energy resolution. The decaying exponential is also evident in the noise distribution of the pixels, detailed in Section 9.2.

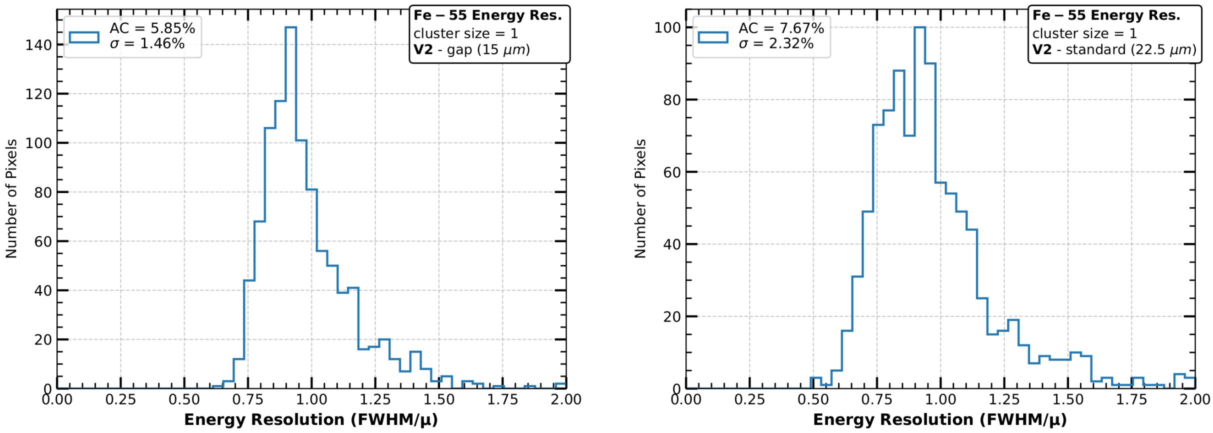


Figure 9.18: (Normalized) Energy resolution (FWHM/μ) distribution of CE-65v2 chip pixels in the Modified with Gap process with a 15 µm pitch (a) and Standard process with a 22.5 µm pitch (b). The unnormalized energy resolution and standard deviation of the distributions are displayed on the top left.

For comparison, Figure 9.18b depicts the equivalent distribution for the 22.5 µm pitch CE-65v2 chip produced in the Standard process. The mean of the distribution is considerably higher at 7.67% and in addition has a higher standard deviation at 2.32%. The decaying exponential is again evident, though it displays a more pronounced tail. As shown in Section 9.2, the 22.5 µm pitch CE-65v2 chip produced in the Standard process does not inherently show a higher ENC noise than its 15 µm pitch Standard process counterpart. The degradation in the measured energy resolution of the Standard process variants is related to the high charge sharing and correspondingly low statistics of the single-pixel cluster spectra. This degradation is also evident in the 15 µm pitch CE-65v2 chip produced in the Standard process.

Dependence on Amplification Scheme

The dependence of the energy resolution on the amplification scheme was also studied by considering the distributions for each of the submatrices in the two CE-65v1 chips. Figure 9.19a depicts the energy resolution distribution of each pixel of the 15 µm pitch CE-65v1 chip produced in the Modified Process. The energy resolution degrades slightly when moving from the AC to the DC submatrix from 7.15% to 8.58%, with both distributions showing a Gaussian with a decaying exponential. The AC distribution has a slightly lower standard deviation at 1.46%, compared to the DC submatrix' 2.14%. The SF submatrix, on the other hand, displays a considerably more smeared distribution with a standard deviation of 2.54% and the worst energy resolution at 9.60%. It is unclear why the SF distribution deviates from its AC and DC counterparts, since the noise distributions from Section 9.2 clearly show the DC submatrix to be the noisiest.

Figure 9.19b depicts the energy resolution of the 25 µm pitch CE-65v1 chip produced in the Standard process. Despite displaying the characteristic Gaussian with the exponential tail, the energy resolution for all three submatrices is worse compared to its Modified counterpart. The AC submatrix shows a mild degradation to 7.82% with a standard deviation of 1.83%. The SF submatrix shows a degradation to 12.84% with a standard deviation of 2.10%, which is around a third higher than its Modified counterpart. The DC submatrix shows the largest degradation to 24.07% with a very large standard deviation at 5.05%. Such a large energy resolution results due to the extreme charge sharing, and associated low charge collection efficiency, evidenced in the DC submatrix of the Standard process CE-65v1, and should be taken with caution.

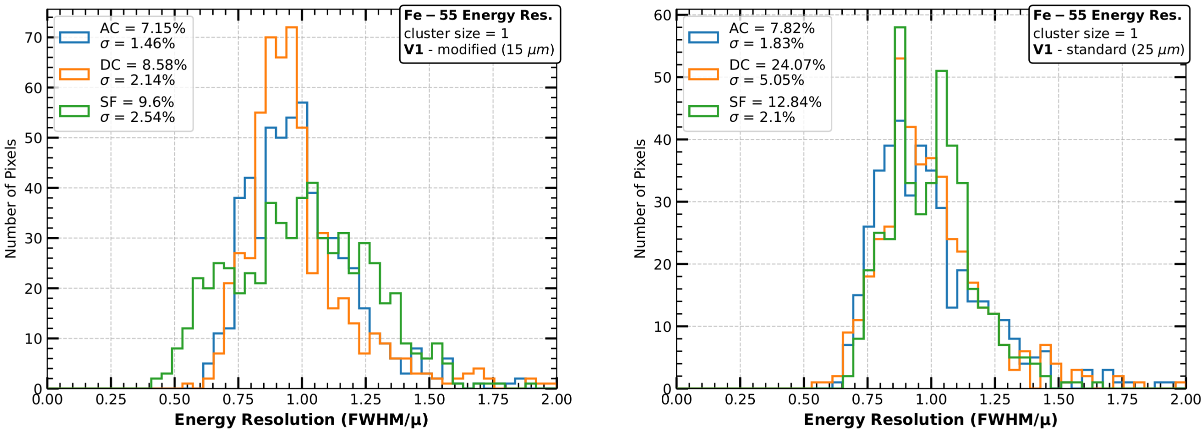


Figure 9.19: (Normalized) Energy resolution (FWHM/μ) distribution of CE-65v1 chip pixels in the Modified process with a 15 µm pitch (a) and Standard process with a 25 µm pitch (b). The unnormalized energy resolution and standard deviation of the three distributions are displayed on the top left.

9.6.3 Pixel-by-Pixel Calibration

The individual pixels of the CE-65v1 and CE-65v2 chips were calibrated by repeating the charge conversion factor computation introduced in Equation 9.5 using the K_α peak positions obtained from the spectra in Figures 9.16a and 9.16b.

Calibration Maps (2-d)

Figure 9.20 summarizes the gain uniformity of the 15 µm CE-65v2 chip in the Modified with Gap process by depicting the K_α peak position for each individual pixel, normalized by the mean K_α peak position for all pixels.

No clear spatial pattern can be discerned, and indeed the standard deviation of the gain between pixels is of O(4%). Similar trends hold for the other three chips, with the variation being of O(4%) in all cases. Figure 9.21 includes for comparison the 2-d distribution of the normalized K_α peak position for the 22.5 µm CE-65v2 chip in the Standard process, where similar conclusions can be drawn.

The gain uniformity of the 15 µm CE-65v1 chip in the Modified process is depicted on Figure 9.22. As for the CE-65v2 chips, no clear spatial pattern can be discerned, although it is evident that the SF submatrix shows a larger variation in the gain. The same is evident for the SF submatrix of the 25 µm CE-65v1 chip in the Standard process, which is depicted on Figure 9.23. The mean K_α peak position across the matrix along with the standard deviation for each of the CE-65v1 and CE-65v2 chips is summarized in Table 9.3.

Gain Variation Distribution (1-d)

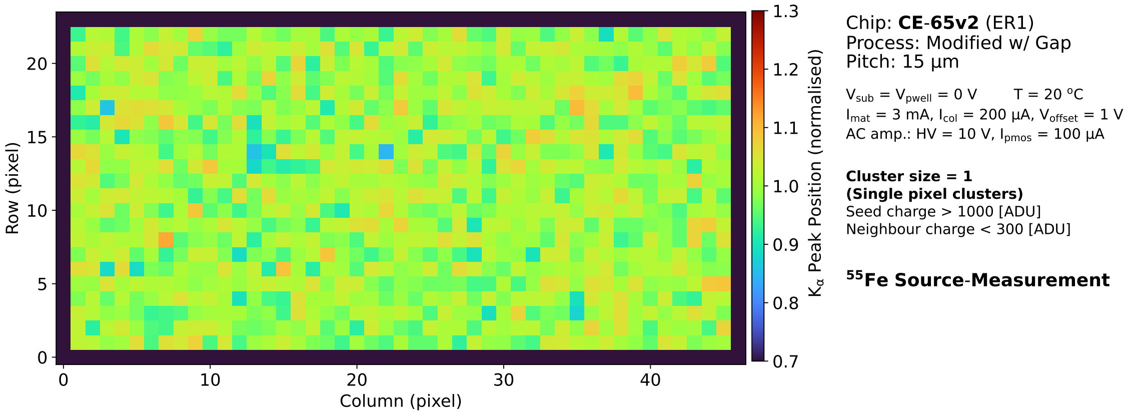


Figure 9.20: Gain distribution of the CE-65v2 chip in the Modified with Gap process with a 15 µm pitch obtained by considering the normalized K_α peak position (Figure 9.16a) for each pixel.

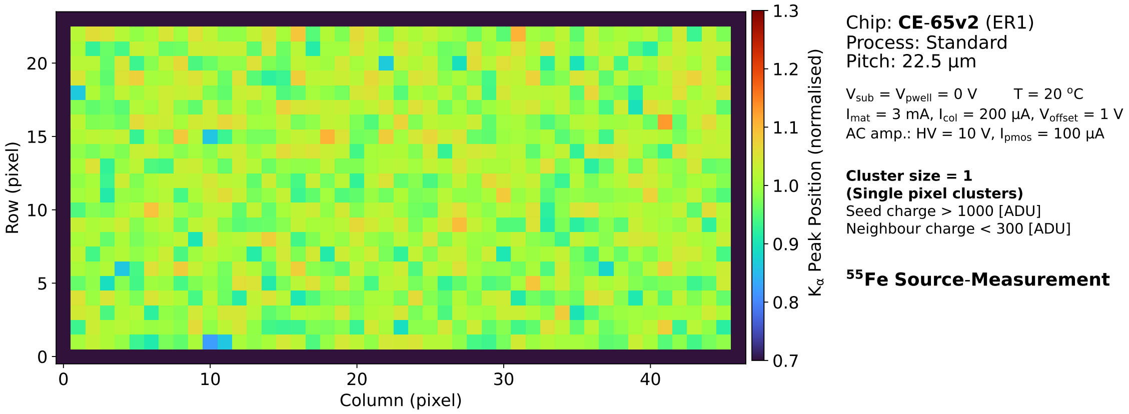


Figure 9.21: Gain distribution of the CE-65v2 chip in the Standard process with a 22.5 µm pitch obtained by considering the normalized K_α peak position (Figure 9.16b) for each pixel.

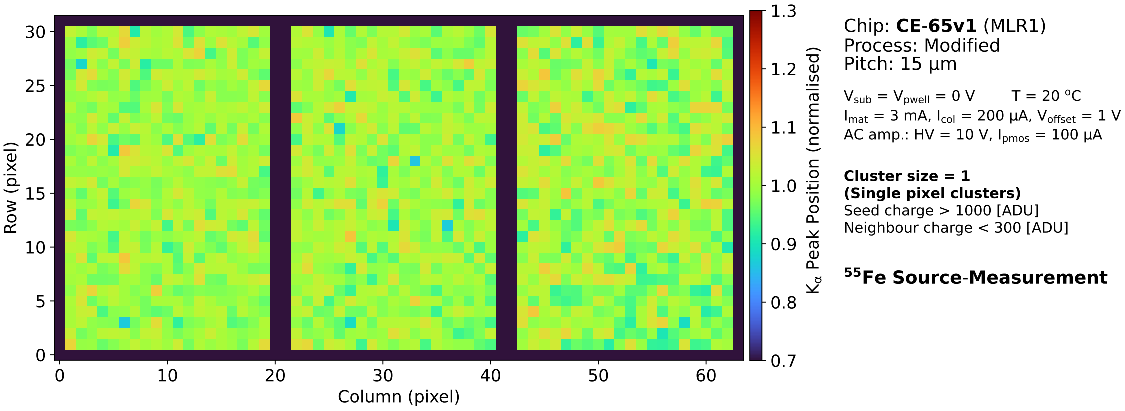


Figure 9.22: Gain distribution of the CE-65v1 chip in the Modified process with a 15 µm pitch obtained by considering the normalized K_α peak position (Figure 9.17a) for each pixel.

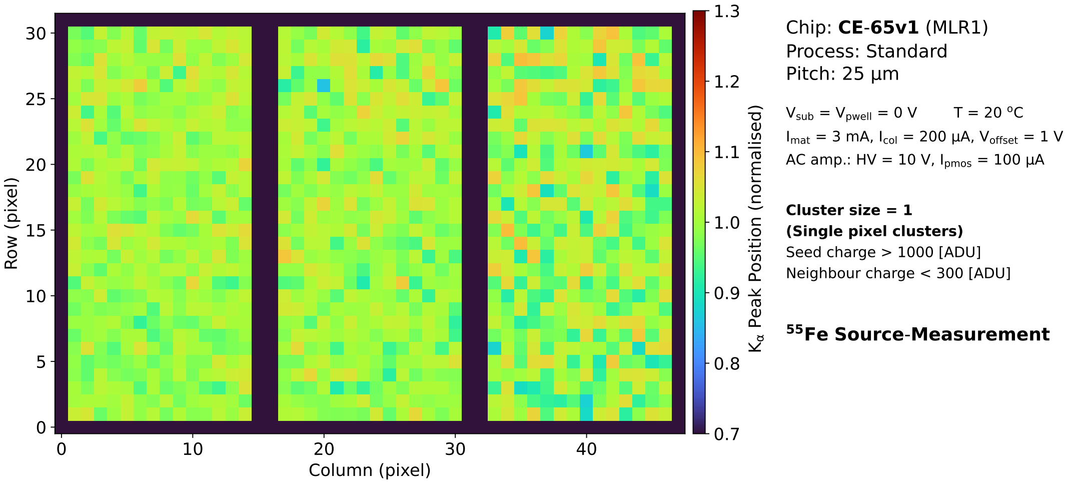


Figure 9.23: Gain distribution of the CE-65v1 chip in the Standard process with a 25 µm pitch obtained by considering the normalized K_α peak position (Figure 9.17b) for each pixel.

Table 9.3: Summary of CE-65v2 and CE-65v1 K_α peak position means in ADUs.

Chip Type	Pitch	Process Mod.	AC-PreAmp.	DC-PreAmp.	Source-Follower
CE-65v2	15 µm	Modified w/ Gap	6863 ± 274	–	–
	22.5 µm	Modified w/ Gap	6990 ± 264	–	–
	15 µm	Standard	6800 ± 274	–	–
	22.5 µm	Standard	6953 ± 277	–	–
CE-65v1	15 µm	Modified	7099 ± 221	5918 ± 196	1840 ± 74
	25 µm	Standard	7292 ± 203	7278 ± 245	2250 ± 114

The 1-d distributions of the normalized K_α peak positions for the 15 µm CE-65v2 Modified with Gap process chip and 22.5 µm CE-65v2 Standard process chip, depicted on Figures 9.24a and 9.24b, respectively, both display a similar mean at $\mu = \sim 6900$ ADUs and standard deviation $\sigma = \sim 270$ ADUs.

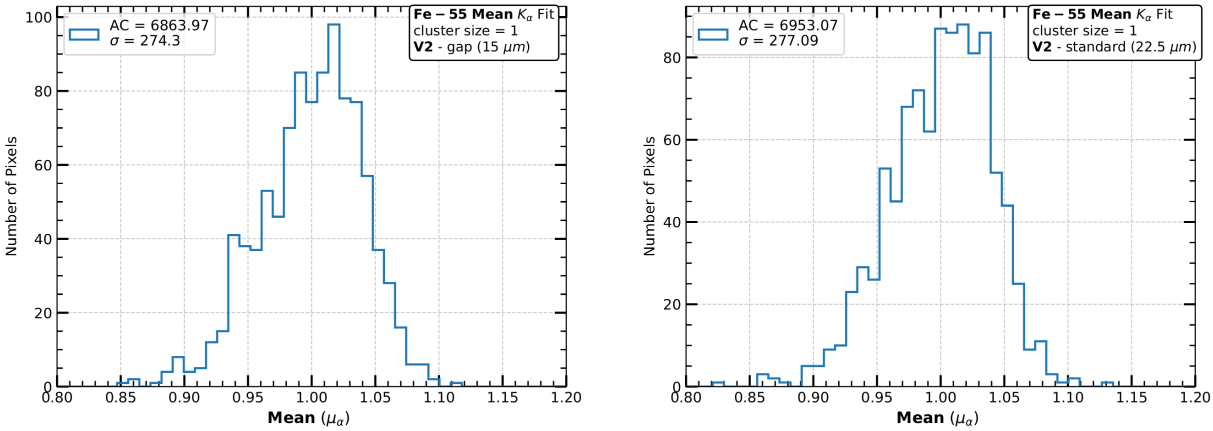


Figure 9.24: (Normalized) K_α peak position (μ) distribution of CE-65v2 chip pixels in the Modified with Gap process with a 15 µm pitch (a) and Standard process with a 22.5 µm pitch (b). The unnormalized mean (μ_α) and standard deviation (σ_α) of the distributions are displayed on the top left.

A priori a symmetric Gaussian is expected corresponding to some central gain value, and smeared by fabrication defects. The precise origin of the one sided tail toward lower K_α peak positions is not known. It could stem from a fabrication defect resulting only in a raising of the sensor capacitance. Although saturation was observed during testbeam studies, it seems an unlikely culprit as the high-mean side of the distribution does not show an abrupt drop, but rather a Gaussian decay. Despite the similarity of the distribution with the ENC distributions given in Section 9.2 when reflected horizontally, no clear correlation between high noise and low gain pixels could be established.

Considering the 1-d distributions of the normalized K_α peak positions for the 15 µm CE-65v1 Modified process chip and its 25 µm Standard process chip counterpart, depicted on Figures 9.25a and 9.25b, respectively, reveals a small dependence of the gain uniformity on the amplification scheme. The AC and DC submatrices display similar Gaussian distributions with a slight tail towards lower gain values. The observed standard deviation is low at O(3%) for the AC and DC submatrices of both the Modified and Standard process chips. By contrast, the SF submatrix shows a less pronounced distribution for the Modified chip, with a larger standard deviation of ∼ 4%, in parity with the CE-65v2 chips. The SF submatrix of the Standard process chip, on the other hand, no longer peaks around a central value, although it still displays a tail towards lower gain. It has the highest standard deviation at ∼ 5%.

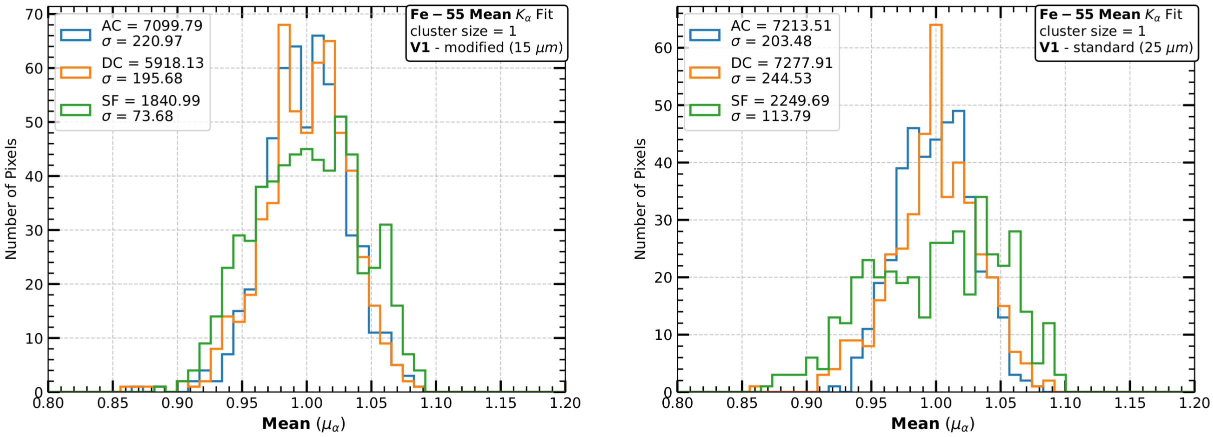


Figure 9.25: (Normalized) K$_\alpha$ peak position (μ) distribution of CE-65v1 chip pixels in the Modified process with a 15 µm pitch (a) and Standard process with a 25 µm pitch (b). The unnormalized mean (μ_α) and standard deviation (σ_α) of the three distributions are displayed on the top left.

Calibrated Fe-55 Spectrum

The signal spectrum for each pixel was considered without imposing any seed or neighbour threshold cuts. The spectra of individual pixels was then added for all pixels. The resulting spectrum, the pixel spectrum, is depicted on Figure 9.26a for the 15 µm Modified with Gap process CE-65v2 chip. The calibrated pixel spectrum, shown in Figure 9.26b, was obtained by using the calibration maps presented in Section 9.6.3 to scale the contributions of each pixel such that

$$s_{ij} \to s'_{ij} = s_{ij}/g_{ij} \tag{9.6}$$

where s_{ij} is the raw signal measured by pixel ($x = i$, $y = j$), and g_{ij} is its normalized gain. In this way, the pixel-to-pixel gain variations can be minimized, and a comparatively clean pixel spectrum can be obtained without suffering from low statistics in the way the single-pixel cluster spectra for individual pixels do.

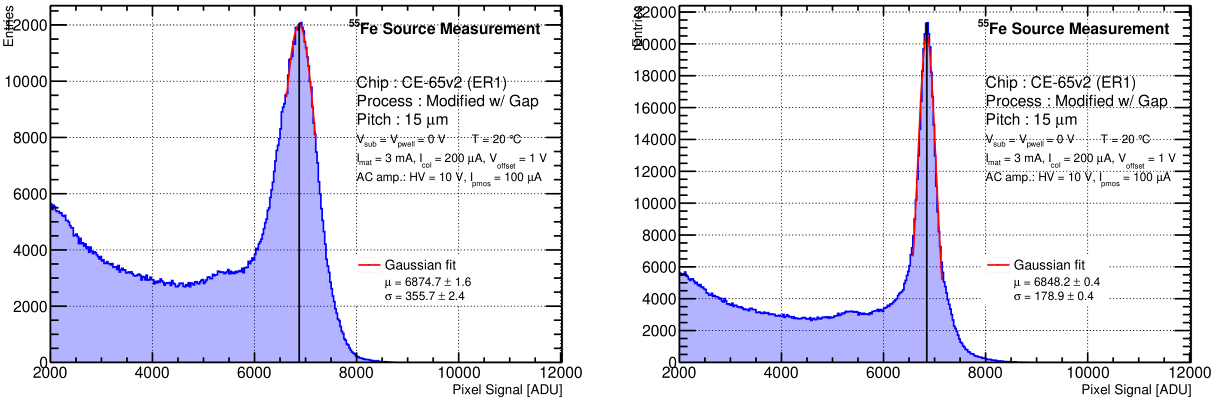


Figure 9.26: The ^{55}Fe spectrum of the pixel signal after CDS of the entire matrix of the CE-65v2 chip in the Modified with Gap process with a 15 µm pitch. The pixel spectrum is shown before (a) and after (b) gain normalisation. The Gaussian fit is marked in red.

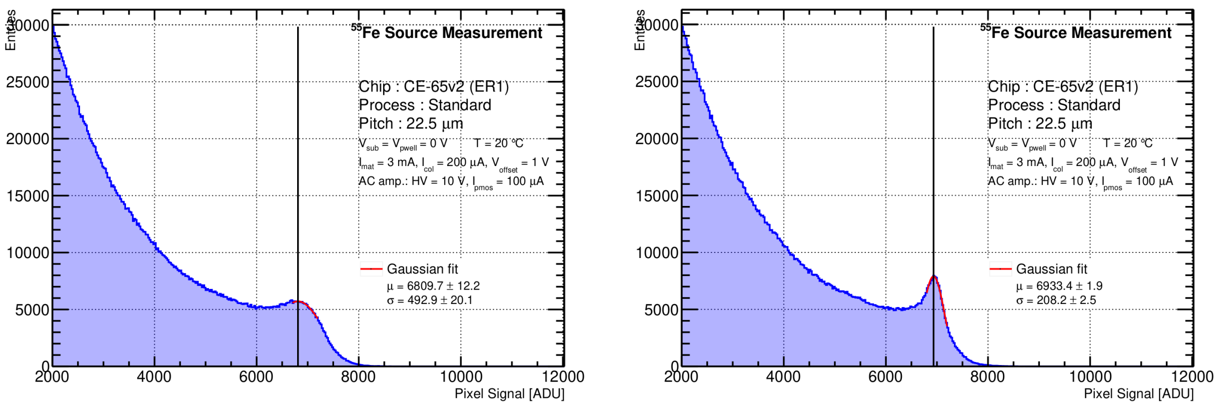


Figure 9.27: The ^{55}Fe spectrum of the pixel signal after CDS of the entire matrix of the CE-65v2 chip in the Standard process with a 22.5 µm pitch. The pixel spectrum is shown before (a) and after (b) gain normalisation. The Gaussian fit is marked in red.

Although the K_β peak is not discernible, the calibration leads to a sizeable reduction in the energy resolution from 12.4% to 6.2%. Figure 9.27a depicts the uncalibrated pixel spectrum for the 22.5 µm Standard process chip, while Figure 10.27b shows the corresponding calibrated result. Here, an even larger improvement in the energy resolution from 17.0% to 7.1% is evident.

9.7 Conclusions

The laboratory characterisation of the CE-65 family of chips confirmed their successful operation in a controlled setting. The study of the sensor response to X-rays shed light on the interplay between amplification scheme, process modification, and pixel pitch: the three *exploration axes* introduced in Section 8.1. The results not only assessed the baseline performance of the CE-65 MAPS prototypes, but also highlighted design-specific strengths and limitations, offering valuable guidance for the development of future chip prototypes using the 65 nm TPSCo CMOS process.

The CE-65v1 and CE-65v2 variants exhibit low-noise performance across different pixel pitches and process modifications. A small dependence on the amplification scheme was determined for the noise performance. However, the majority of the pixels across all chip variants displayed an ENC in the 15 e^- - 25 e^- range.

Gain calibration studies demonstrated a high gain uniformity of O(4%) for all considered chips. Nominal gain values showed little dependence on pixel pitch and process modification. The relative gain of the three amplification schemes (SF:DC:AC) was roughly 1:4:3 for the 25 µm Standard process CE-65v1 chip and 1:3:4 for its 15 µm Modified process counterpart. This is in good agreement with simulation where an increased gain of O(4) is expected due to the amplification.

The Standard process chips displayed significant charge sharing, with ^{55}Fe spectra being dominated by cluster sizes of 4 or more. This correspondingly resulted in a decreased energy resolution. In contrast, the Modified and Modified with Gap variants exhibited lower charge sharing, with ^{55}Fe spectra dominated by single- and two-pixel clusters.

Among the amplification schemes, the AC-coupled preamplifier achieved the highest charge collection efficiency (∼ 97%) and best energy resolution (∼ 7%). The application of a large biasing voltage (10 V) allowed for the full depletion of the epitaxial layer. Initial laboratory studies of the CE-65v1 chip were instrumental in identifying the AC submatrix as the most performant, and motivating its selection as the sole amplification scheme in the CE-65v2 chip.

The validation and characterisation of the CE-65 family of chips in a laboratory setting served as a critical stepping stone to the testbeam studies of the CE-65v2 chip, which explored all four exploration axes from Section 8.1: amplification scheme, process modification, pitch variation and matrix geometry to optimize spatial resolution and detection efficiency.

10 Testbeam Analysis

The characterisation of novel sensor technologies typically rests on three pillars: simulation studies, laboratory tests, and testbeams. Following the initial study of sensor properties in a controlled lab setting, where basic functionality of the chip is verified, testbeam campaigns represent a critical next step in assessing real-world performance. Testbeam campaigns offer the opportunity to study the response of a sensor to MIPs, providing realistic performance benchmarks for metrics crucial to the applications of the sensors in HEP, including the hit-detection efficiency and spatial resolution. In contrast to lab tests, where one typically relies on a radioactive source with essentially no tracking information, testbeams leverage beam telescopes to precisely reconstruct particle trajectories, allowing the determination of the these key metrics. These metrics underline the suitability of the 65 nm TPSCo CMOS process in high-resolution vertexing applications, with the detection efficiency and spatial resolution serving as the primary arbiters for the design choices.

While Chapter 9 presents the lab characterisation of both the CE-65v1 and CE-65v2 chips, this chapter focuses solely on the testbeam campaigns of the CE-65v2 chip. This chip adopts the best-performing amplification scheme from CE-65v1, and introduces additional design variations, notably the staggered matrix geometry. The chapter begins with a brief introduction to testbeam analysis and the adopted analysis framework. The CE-65v2 testbeam at the Super Proton Synchrotron (SPS) [215] at CERN is detailed, with the main results including the obtainable hit-detection efficiency, spatial resolution, and the cluster properties as a function of process modifications and pitch. Subsequently, the CE-65v2 testbeam at DESY [216] is presented, where the cluster properties with respect to biasing voltage were explored. Finally, the chapter concludes with a summary of the main findings and their implications for future sensor development.

Parts of the SPS testbeam results covered in this chapter have been published in Refs. [4, 196]. Similar phrasing was used for relevant sections.

10.1 Methodology

The core task during a testbeam is the characterisation of the performance of a given sensor by studying its response to MIPs at a beam facility. Testbeam campaigns offer the opportunity to evaluate a sensor under real-world operating conditions, unlike the laboratory measurements presented in Chapter 9. This is typically done by arranging the device under testing (DUT) into a beam telescope consisting of several reference detectors whose properties have been well established. Figure 10.4 depicts the adopted setup during the CE-65v2 testbeam campaign at the SPS. The reference detectors are arranged so that the trajectories of the particles passing through the DUT can be determined with utmost precision. By combining multiple measurements of the same particle as it traverses the reference detectors, a track can be reconstructed with a resolution that is significantly better than the spatial resolution of the individual reference detectors. The interpolated track position on the DUT can be compared with the measured sensor response in order to determine a number of core metrics.

10.1.1 Performance Metrics

The first of these is the *hit-detection efficiency.* It can be understood as the probability with which a given particle track elicits a discernible response in the DUT. Formally, it is defined as the ratio between the number of tracks with an associated cluster k, and the total number of tracks N:

$$\epsilon = \frac{k}{N} \tag{10.1}$$

The details of how tracks are associated to clusters are given in Section 10.1.2.

The second metric, the *residual*, is intimately related to the hit-detection efficiency. For a track with an associated cluster, the residual in the x, y dimension is defined as the distance between the track intercept on the DUT plane, and cluster centre, as determined by a clusterisation method:

$$\Delta x = x^{*} - x_{cl.} \tag{10.2}$$

$$\Delta y = y^{*} - y_{cl.} \tag{10.3}$$

where the x^{*} denotes the interpolated position, and the $x_{cl.}$ the cluster centre. Throughout the testbeam campaigns detailed in this chapter the mean of the residuals in the x and y dimensions, Δd, was used to estimate the intrinsic spatial resolution as

$$\sigma_{DUT} = \sqrt{(\Delta d)^2 - \sigma_{tel.}^2} \tag{10.4}$$

where the $\sigma_{tel.}$ denotes the telescope resolution, which depends on both particle momentum and type.

Many of the design levers of the CE-65v2 chip revolve around the tuning of cluster properties, including the cluster size and shape, in order to optimize the aforementioned metrics. It is thus worthwhile to introduce a third metric, the *accumulated charge ratio*, which was used to quantify the observed charge sharing in Sections 10.2.5 and 10.3.2. The accumulated ratio is defined by summing the charge of individual pixels in a cluster successively, until a cluster size of 9 is reached. The accumulated ratio is then normalized by the total cluster charge, calculated as the sum of the charge of all pixels in a 3×3 window around the seed pixel:

$$f_{acc.}^{i} = \frac{\sum_{j=1}^{i} Q_j}{\sum_{j=1}^{9} Q_j} \tag{10.5}$$

where $f_{acc.}^{i}$ denotes the accumulated charge ratio for the i^{th} pixel and Q_i denotes its charge. The indices in the computation of the accumulated charge ratio are assigned by charge ordering, with the first pixel being the most energetic.

10.1.2 Corryvreckan

The Corryvreckan software framework [3, 217] provides a toolset for the reconstruction and analysis of testbeam data. The framework spans the entire pipeline of a typical testbeam analysis, from the reading of the raw detector data to the final observables in the form of efficiencies and residuals. The framework is organised in individual user-configured *modules*, allowing extendability, which is crucial for the framework to support the wide variety of sensors, geometries, and analyses that can be associated with a generic testbeam campaign. The modular nature of the framework allows additions and modifications of the analysis chain without necessitating large rewrites of the source code. In the case of the CE-65v2 chip, this allowed the extension of the current analysis module, and the implementation of the staggered matrix geometry. The adopted analysis chain for the CE-65v2 testbeam at SPS is depicted on Figure 10.1, where each individual module in the DUT analysis step can be seen.

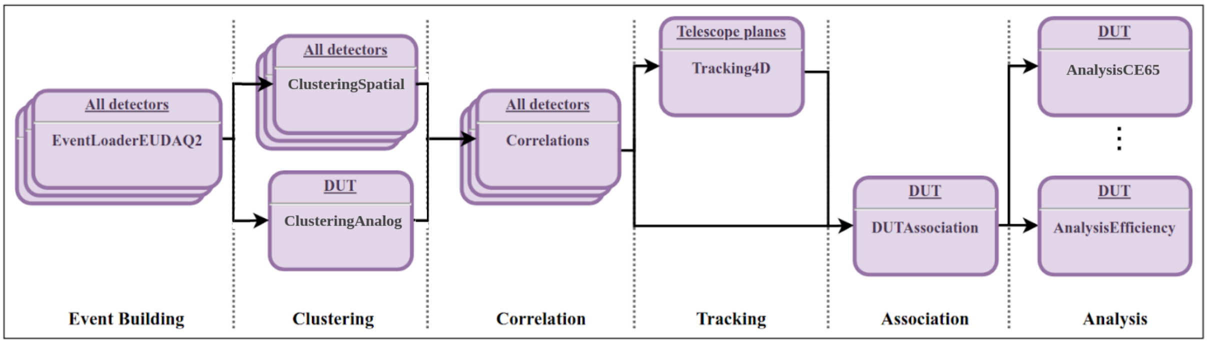

All detectors
EventLoaderEUDAQ2
All detectors
ClusteringSpatial
DUT
ClusteringAnalog
All detectors
Correlations
Telescope planes
Tracking4D
DUT
DUTAssociation
DUT
AnalysisCE65
DUT
AnalysisEfficiency
Event Building
Clustering
Correlation
Tracking
Association
Analysis

Figure 10.1: Overview of the Corryvreckan analysis pipeline for the CE-65v2 testbeam campaigns. Individual modules are illustrated as purple boxes, while the flow of data is indicated by black arrows. Figure adapted from Ref. [3].

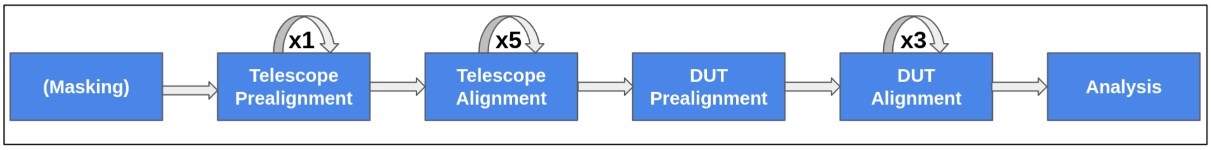

(Masking)
x1
Telescope Prealignment
x5
Telescope Alignment
DUT Prealignment
x3
DUT Alignment
Analysis

Figure 10.2: Schematic of the testbeam analysis steps. Each step consists of reconstructing the testbeam data, and applying the appropriate modules to process the data, as depicted for the final Analysis step in Figure 10.1. The (pre-)alignment steps produce updated geometry files detailing the telescope, which are sequentially propagated though the chain until the final Analysis step.

Telescope (Pre-)Alignment

The first step in the Corryvreckan reconstruction chain, depicted on Figure 10.2, is the alignment of the telescope. A precise alignment of the beam telescope is crucial, as minor offsets of a fraction of a millimetre can be enough to shift the computed position by several pixels, significantly impacting subsequent analysis. The physical offset in the z-axis along the beam direction must be measured physically and provided to the Corryvreckan framework. The alignment procedure is segmented in two separate steps, whereby the detector planes are first pre-aligned with respect to a reference detector, and subsequently precisely aligned by a dedicated *alignment* module. In order to avoid biasing the analysis, the DUT is excluded from all steps during the pre-alignment and alignment of the beam telescope.

The goal of the pre-alignment step is to provide a coarse description of the telescope geometry for later fine-tuning. Pre-alignment begins with the computation of the spatial correlations between detector planes, defined as the distance any cluster on the given detector plane and any cluster on the reference detector. During the analysis of the CE-65v2 data, the first ALPIDE [185] from the downstream arm was chosen as the reference detector. The other telescope planes are then translated in the x-y plane with the goal of centring the correlation peaks for each of the planes. Rotations are not performed during the pre-alignment step, and instead only take place during alignment.

Once a coarse estimate of the telescope geometry has been performed by the *prealignment* module, the description is passed to the *alignment* module where an iterative minimisation procedure is performed on the computed tracks. In particular, while pre-alignment step did not require an explicit reconstruction of the tracks, as shown in Figure 10.1, track reconstruction is a prerequisite step of the precise alignment of the beam telescope.

Track reconstruction begins with the identification of clusters on all detector planes during a given event. Track candidates are then built by considering a straight line interpolation between two given clusters on the first and last detector planes of the beam telescope. Clusters on all remaining detector planes along the track candidates are looked for in a user-defined elliptical search window of 100 µm × 100 µm. If a cluster is found on a subsequent detector plane within the search window, then the cluster is added to the track candidate, and the track is refitted. For the SPS testbeam the high beam energy ensured minimal multiple scattering, and thus a straight-line model sufficed for track construction. At the DESY testbeam the considerably higher multiple scattering motivated the use of the General Broken Lines (GBL) algorithm [218, 219], whereby tracks are allowed a kink angle at every detector plane to scattering effects. If a cluster can be associated on every detector plane for the given track model, then the track candidate becomes a reconstructed track.

The precise goal of the alignment procedure is the minimisation of the χ^2 of all computed tracks, via the iterative translation and rotation of the detector planes with respect to the reference detector. Once the minimisation algorithm converges, the telescope is considered to be aligned and its geometry is frozen for subsequent steps[1].

DUT (Pre-)Alignment

Following the alignment of the telescope the precise position of the DUT with respect to the reference plane must be determined. This proceeds similarly to the alignment of the telescope, with an initial coarse pre-alignment, an a subsequent fine-tuning by the *AlignmentDUTResidual* module. During the pre-alignment the correlations of the DUT with the reference plane are minimised by translations on the x-y plane. This is followed by the alignment step, where rather than minimising the track χ^2 as was done for the telescope alignment, the DUT plane is iteratively translated and rotated to minimise the unbiased residuals between the interpolated track position on the DUT and its associated cluster. The tracks are constructed as described during the telescope alignment step, with an additional track quality cut of $\chi^2/n_{dof} < 2$ to prevent misalignment due to poor quality tracks. The DUT clusters are associated similarly to in the track fitting step, by considering an initial elliptical search window of 100 µm × 100 µm. In the case where multiple clusters are found, the combined contribution of all residuals is considered for that track during the minimisation of the residuals.

---

[1] While the telescope geometry is frozen for analysis steps, it is generally good practice to repeat the alignment if there have been appreciable changes to the telescope geometry, e.g. thermal expansion during long runs.

DUT Analysis

The final step in the Corryvreckan analysis pipeline is the computation of the performance metrics. In the case of the CE-65v2 chip this entails primarily the hit-detection efficiency and the x,y-residuals. In the Corryvreckan framework this is performed by the *AnalysisEfficiency* and *AnalysisCE65* modules. The efficiency was computed as the number of tracks with associated clusters divided by the total number of tracks, as defined in Equation 10.1. The uncertainty is given by the Clopper-Pearson [220] confidence interval of 1 σ. The residuals were extracted via the *AnalysisCE65* module, which is a derived class of the *AnalysisDUT* module, where the x, y residuals are defined by Equation 10.2.

Coordinate Systems

The Corryvreckan framework relies on three basic coordinate systems for the internal representation of a given testbeam telescope. These can be sub-divided into two *local* coordinate systems, and one *global*. The first local coordinate system represents the actual detector channels, and thus for the square and staggered geometries adopted in this work, represents the row and column of a given hit. It corresponds closely to the raw position of where the hit was recorded. By definition, this coordinate system is centered at the centre of the lower left pixel of the given detector. For the square matrix arrangement, this coordinate systems already defines a cartesian coordinate space supporting vector addition. This, in principle, allows for the computation of the cluster centre, and other distance metrics relevant to the analysis pipeline. However, for more non-trivial geometries, including the staggered matrix arrangement, depicted on the left of Figure 10.3, this is no longer the case. Instead, the second local coordinate system is introduced, which is explicitly a right-handed cartesian coordinate system centred at the geometric centre of the given detector plane. It is typically referred to as the *local* coordinate system, and appears in the centre of Figure 10.3. A set of helper functions are defined to translate the position in the first local coordinate system to the local cartesian coordinate system, which for the square matrix take the trivial form of

$$\begin{aligned} x_{local} &= p_x \cdot \left(x_{col.} - \frac{N_{col.}-1}{2} \right) \\ y_{local} &= p_y \cdot \left(y_{row} - \frac{N_{row}-1}{2} \right) \end{aligned} \tag{10.6}$$

where the p_x, p_y denote the pixel pitches in their respective dimensions and $N_{col.}$, N_{row} denotes the number of columns and rows, respectively. The translation of the y coordinate has to be modified for the staggered matrix arrangement, yielding

$$y_{local} = p_y \cdot \left(y_{row.} - \frac{N_{row}-1}{2} \pm \frac{f_{stag.}}{2} \right) \tag{10.7}$$

where $f_{stag.}$ represents the pixel fraction by which the matrix is staggered, which in the case of the CE-65v2 chip is $f_{stag.} = \frac{1}{2}$. The addition or subtraction of the additional pixel fraction follows the opposite parity of the staggering for the given column, i.e. if column x_i is staggered positively with respect to the geometric centre, then the minus sign is taken during the computation of y_{local}.

While the local coordinate systems allows for highly simplified calculations on each given detector plane, it is necessary to introduce a third coordinate system, the *global* coordinate system, which relates all detector planes to each other. In particular, the global coordinate system is defined with respect to the beam axis, with the beam direction defining the z-axis, as depicted on the right of Figure 10.3. The origin of the z-axis is defined by the user-specified measurements of the detector planes. In the case of the CE-65v2 analyses, this was chosen to be the first telescope plane from the beam direction.

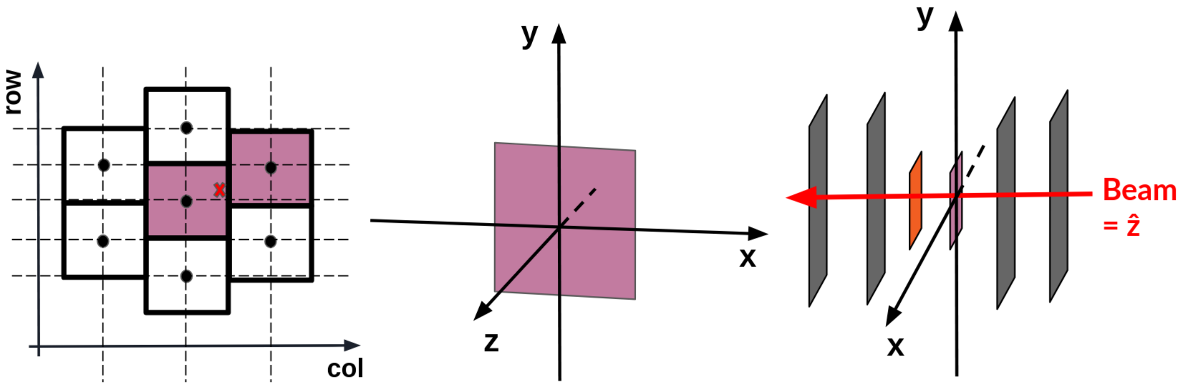


Figure 10.3: Illustrations of the three coordinate systems in the Corryvreckan framework. The first local coordinate system (a) describes positions on the given detector plane with respect to the detector channels, and is not necessarily Cartesian. It is typically centered at the corner of the given detector. The second local coordinate system (b) is a Cartesian coordinate system centered at the geometric center of the detector plane, depicted in pink. The global coordinate system (c) relates the local Cartesian coordinate systems to each other by defining the z-axis along the beam direction. The DUT and trigger planes are marked in pink and orange, respectively.

10.2 CERN SPS 2024 Testbeam

In April 2024 a testbeam campaign for the CE-65v2 chip was conducted at CERN SPS [215]. The testbeam focused primarily on evaluating the achievable spatial resolution and hit-detection efficiency of the different sensor variants using high energy beams. This section presents the testbeam setup, followed by a detailed description of measured efficiencies, spatial resolution, and charge sharing results.

10.2.1 Setup

Testbeam Facility

Test beam measurements were conducted at CERN SPS's H6 beamline, part of CERN's North Area which houses a number of smaller experiments and beamlines dedicated to detector development. The area is fed by a slowly extracted proton beam from the SPS at momenta of up to 400 GeV/c. The extracted proton beam is steered into a transfer line where a number of targets sit: T2, T4, and T6. Splitter magnets divide the beam among the different targets, with the T4 target feeding the H8 and H6 beamlines, where the CE-65v2 testbeam was conducted. The T4 target consists of Beryllium, which when struck produces a broad spectrum of secondary particles, including pions, kaons, protons, and electrons. A system of dipole magnets, denoted the wobbling station, allows for control over the production angle and charge sign for each beamline. In the case of H6, the beam is always produced at zero angle, limiting the flexibility in tuning but simplifying operation. After production, the beam is collimated and transported to the H6 beamline, where it serves a versatile facility for detector characterisation, offering a broad range of particle types and momenta of up to 200 GeV/c. For all results presented in this thesis, a 120 GeV/c mixed hadron beam was used.

Beam Telescope

A telescope consisting of six ALPIDE [185] planes, serving as reference detectors, and a DPTS [198] chip acting as the trigger, was used during the testbeam measurements. The ALPIDE reference detectors comprise 1024 × 512 pixels with a pitch of 29.24 µm × 26.88 µm. With its binary readout, each ALPIDE chip achieves a spatial resolution of 5 µm. The ALPIDE chips were mounted on metallic holders fixed to a perforated breadboard, as shown in Figure 10.4. Three of the telescope planes were positioned upstream of the DUT, while the remaining three were placed downstream, with 25 mm spacing between them. The DPTS chip was mounted on a movable stage, to allow alignment of its sensitive area with that of the DUT. The telescope resolution was estimated to be 2.2 µm using a telescope optimizer tool [221] based on the adopted geometry. Cooling pipes connected the DUT holder to a water chiller in order to maintain the DUT at a constant temperature of 20°C throughout the measurements reported in this chapter.

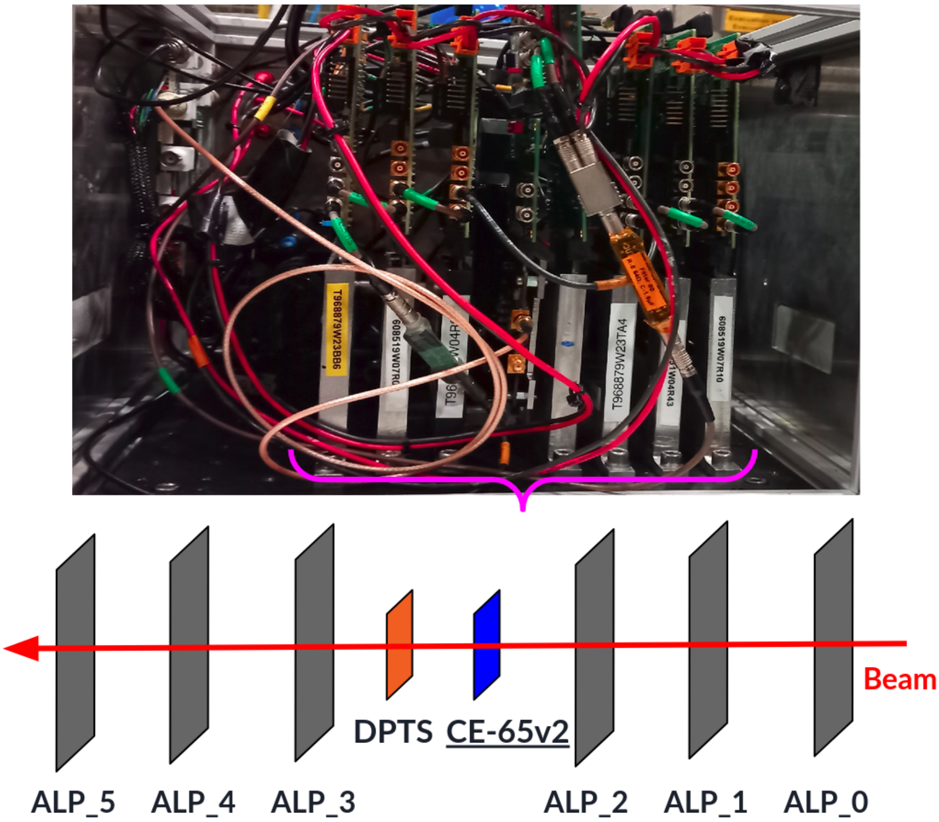


Figure 10.4: Photograph (top) and schematic (bottom) of the CE-65v2 telescope at the SPS testbeam. The DUT can be identified by the blue marking on the unlabelled metallic holder. During data taking the setup was covered to prevent light exposure of the detector planes.

10.2.2 Analysis Strategy for SPS

Data analysis was carried out using the Corryvreckan framework introduced in Section 10.1.2. Figure 10.5 depicts the hitmaps for each of the ALPIDE planes, where the outline of the trigger shadow, corresponding to a higher number of hits, can clearly be discerned.

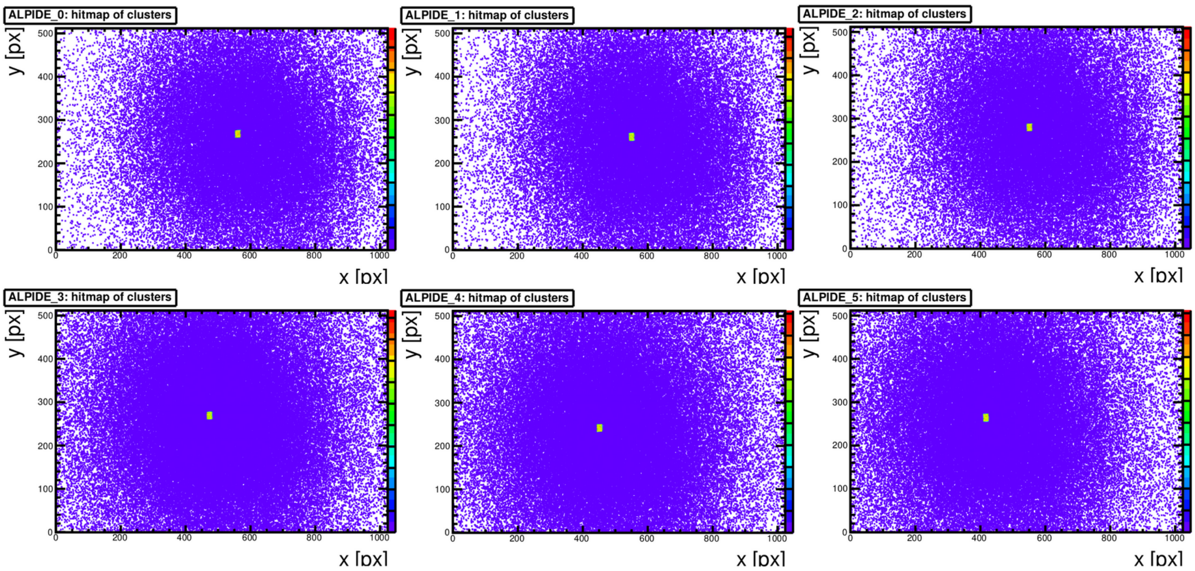


Figure 10.5: Fig 11.5 Hitmaps of the six ALPIDE planes during an SPS testbeam run. The DPTS trigger shadow can be discerned as the small yellow square. Plane labelling follows the convention used in Figure 10.4.

Cluster Position Reconstruction

Clusterisation was performed using two methods: *window* and *cluster*. The window method is equivalent to the matrix method introduced in Section 9.3.2, but is referred to as window throughout this chapter in accordance with Corryvreckan terminology. Clusters on the CE-65v2 chip were built by summing all pixels in a 3 × 3 window around a seed pixel passing an 100 e^- threshold. The maximum cluster size is thus by definition 9. The window method was selected as the default clusterisation method as it allows the evaluation of the full potential of the process technology, without making assumptions about a given readout. Additionally, as it makes no reference to neighbouring thresholds, the window method is ideal for the study of cluster properties, including the accumulated charge ratio computed in Sections 10.2.5 and 10.3.2.

The cluster method, on the other hand, corresponds to the size-n cluster method also introduced in Section 9.3.2, whereby neighbours are added to a cluster by recursively checking if adjacent neighbours pass a given threshold. The cluster method allowed a more realistic evaluation of sensor performance. Threshold scans, expressed in units of electrons and hereafter referred to as electron threshold scans, were performed mimicking noise reduction procedures as implemented in chips used in active detector systems. For the electron threshold scans, identical seed and neighbour threshold cuts were applied for the pixels in the cluster.

For both clusterisation methods, the cluster centre was defined as the centre-of-mass position of all pixels in a cluster

$$(x_{cl.}, y_{cl.}) = \frac{\sum_{i\in cl.}(x_i, y_i)\cdot q_i}{\sum_{i\in cl.} q_i} \tag{10.8}$$

where $(x_i,\, y_i)$ denotes the position vector of pixel i in the cluster, and q_i denotes its charge. For the digitisation tests performed in Section 10.2.4, the charge q_i was digitised to different levels of precision. For all other sections in this chapter, the full analog information was used.

In order to prevent the formation of clusters due to noise fluctuations, particularly at low electron thresholds, a signal-to-noise ratio greater than 3 was required for all seed pixels. The noise was computed as the standard deviation of the Gaussian fit to the CDS distribution for each pixel, as detailed in Section 9.2. Multiple cluster candidates were considered corresponding to each pixel passing the required seed e$^-$ threshold and signal-to-noise ratio. Clusters within a 75 µm radius were associated with a given track as described in Section 10.1.2, corresponding to just over 3-4 times the pixel pitch. The nearest associated cluster was selected in cases of multiple candidates.

Charge Spectra

The charge deposited in the epitaxial layer of a silicon sensor is governed largely by its thickness, since the energy loss per unit length is comparable across different thicknesses, as illustrated by Figure 7.4b introduced in Section 7.2. At low silicon thickness of order $\mathcal{O}(100)$ µm or less, the combined effect of the detector noise and resolution along with the non-negligible electron binding energy broadens the distribution considerably. The observed straggling function can be pragmatically described by the convolution of a Landau and a Gaussian [180]:

$$f(x) = \frac{A}{\xi\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} f_L(\tau; x_{MPV}, \xi)\cdot e^{-\frac{(x-\tau)^2}{2\sigma^2}}\, d\tau \tag{10.9}$$

where the four fit parameters are the area under the curve (A), the MPV (x_{MPV}) and the scale (ξ) of the Landau distribution, and finally the width (σ) of the Gaussian function. Figure 10.6 depicts the normalized distributions of the cluster charge in electrons for the 15 µm and 22.5 µm Modified with Gap and Standard process CE-65v2 prototypes measured at the SPS testbeam. All four distributions were fit with a Landau-Gaussian convolution, which describes the observed energy loss well. As expected, the cluster charge is virtually identical for all four chip variants, as these all share the same sensor thickness. Although the charge collected by a typical seed pixel differs greatly between the different variants, across the entirety of the cluster the charge is very similar for all variants.

Figure 10.7 depicts the normalized distributions of the seed charge in electrons for all four CE-65v2 variants. The MPV of the seed energy of all four chips lies considerably below the MPV of the cluster charge, which is explicitly marked by the narrow grey band. A clear trend emerges of a high energy fraction being carried by the seed pixel in the case of the Modified with Gap chips, and a small fraction being carried by the seed pixel in the case of the Standard variants. This diverging behaviour manifests as a design trade-off, where the large energy fraction carried in the Modified with Gap chips enables operation at higher electron thresholds, increasing robustness and lowering front-end complexity and data read-out challenges. On the other hand the large charge sharing of the Standard process enables a more precise reconstruction of the track position if information encoded in the charge carried by neighbouring pixels is used. The extent to which

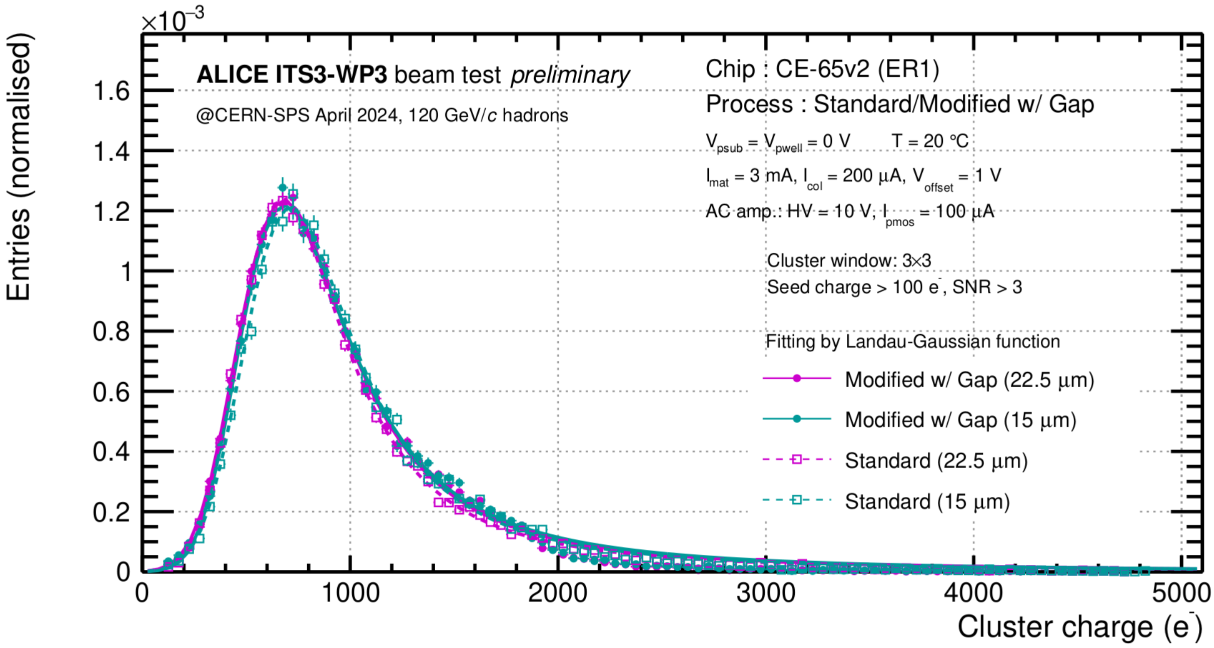


Figure 10.6: Normalised distribution of total cluster charge at a seed threshold of 100 e^- for the 4 considered CE-65v2 chips. All four distributions were fit by a Landau-Gaussian convolution, as defined in Equation 10.9. The window method was used for reconstruction.

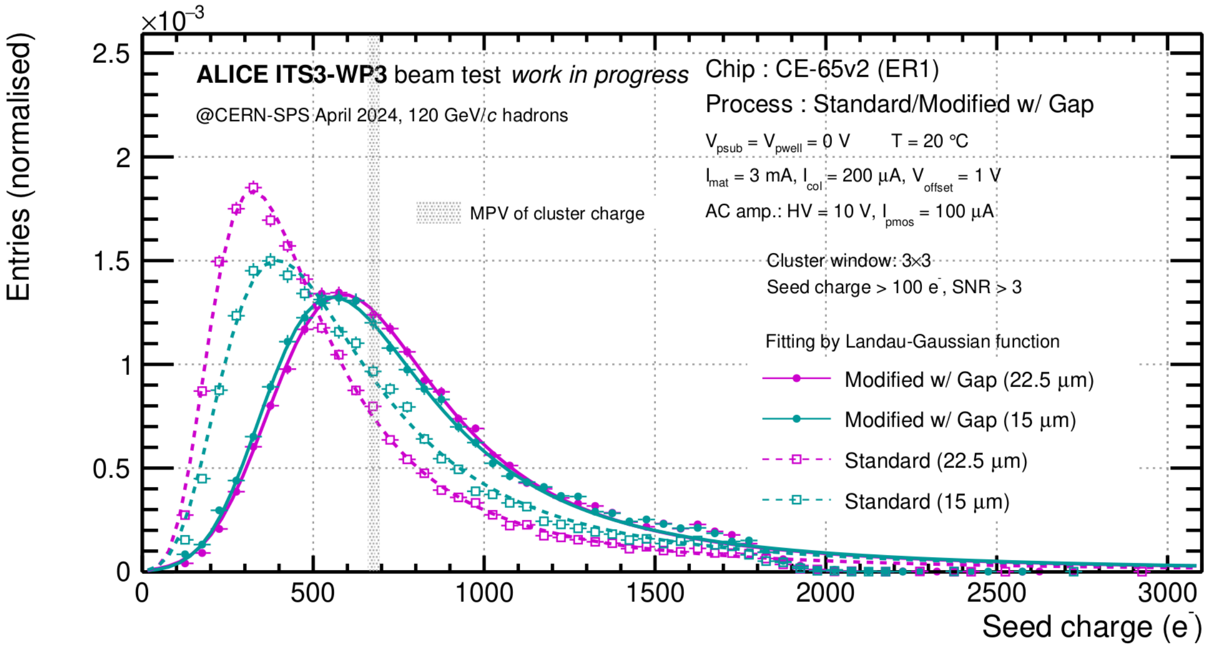


Figure 10.7: Normalised distribution of seed charge at a seed threshold of 100 e^- for the 4 considered CE-65v2 chips. All four distributions were fit by a Landau-Gaussian convolution, as defined in Equation 10.9. The MPV of the cluster charge from Figure 10.6 is overlaid for reference. The window method was used for reconstruction.

this is possible with different levels of digitisation, and thus complexity, is explored in Section 10.2.4.

The Modified with Gap chips, in particular, both have a seed energy MPV of ∼ 600 e$^-$, which is ∼ 14% lower than that of the cluster charge. There is little dependence on the pitch for the MPV of the seed energy, which reflects the charge sharing studies detailed in Section 10.2.5. This effect, however, is much more pronounced for the Standard process, where the 22.5 µm shows a seed energy MPV that is ∼ 55% lower than that of the cluster charge, highlighting the large sharing of the Standard process. The 15 µm variant shows a seed energy MPV that is ∼ 43% lower than that of the cluster charge, which although lower than its 22.5 µm counterpart, still shows substantial charge sharing.

At around ∼ 1700 e$^-$, the seed energy distributions of the Modified with Gap chips exhibit a localized excess inconsistent with the expected Landau-Gaussian fit. The corresponding cluster energy distributions show a similar mild excess in this range, followed by a sharp drop at approximately ∼ 1900 e$^-$. This behaviour suggests pixel saturation above a certain signal threshold, with the transition smeared by pixel capacitance. As a result, the saturation manifests as a bump in the seed energy distribution rather than a sharp cutoff. While this bump is much less pronounced in the Standard process variants, it is still evident. Investigations into the origin of this bump, and the abrupt drop in signal, are ongoing at the time of writing.

10.2.3 Hit Detection Efficiency

The hit-detection efficiency of the four chip variants was studied as a function of electron threshold using the cluster method to quantify the trade-off between detection performance and noise robustness at higher operating thresholds. Figure 10.8 depicts the efficiency as a function of electron threshold for the Modified with Gap process chip variants. An efficiency of over 99% is achieved up to ∼ 180 e$^-$ for both the 22.5 µm and 15 µm chips. The efficiency declines as the electron threshold is increased. Given the relatively large association window of 75 µm, this trend is expected irrespective of the cluster size, as even a single-pixel cluster is sufficient to detect a track impinging upon any point of the pixel. In practical terms, the efficiency begins to decrease as the threshold excludes clusters associated with low charge depositions, corresponding to the lower end of the distribution shown in Figure 10.6. The difference in the measured detection efficiency with respect to pitch size is marginal, with the 22.5 µm pitch maintaining a slightly higher efficiency.

Figure 10.9 depicts the efficiency as a function of electron threshold for the Standard process chips. By comparison, these chips achieve an efficiency of over 99% up to ∼ 130 e$^-$ and ∼ 150 e$^-$ for the 22.5 µm and 15 µm chips, respectively. The decrease in efficiency is considerably faster than for the Modified with Gap chips, with the Standard process 22.5 µm pitch chip exhibiting a steeper drop than its 15 µm counterpart. This behaviour results from the much smaller fraction of cluster charge carried by the seed pixel. The disparity is clearly visible in Figure 10.7, where a threshold of 200 e$^-$ excludes a much larger fraction of events in the Standard process chips.

In-pixel Efficiency

The large matrix of the CE-65v2 chip allowed dedicated high-statistics runs to be performed in order to study the in-pixel distributions of the efficiency, and the resolution, for each of the four chip variants presented. A threshold of ∼ 170 e$^-$ was chosen in order to demonstrate the

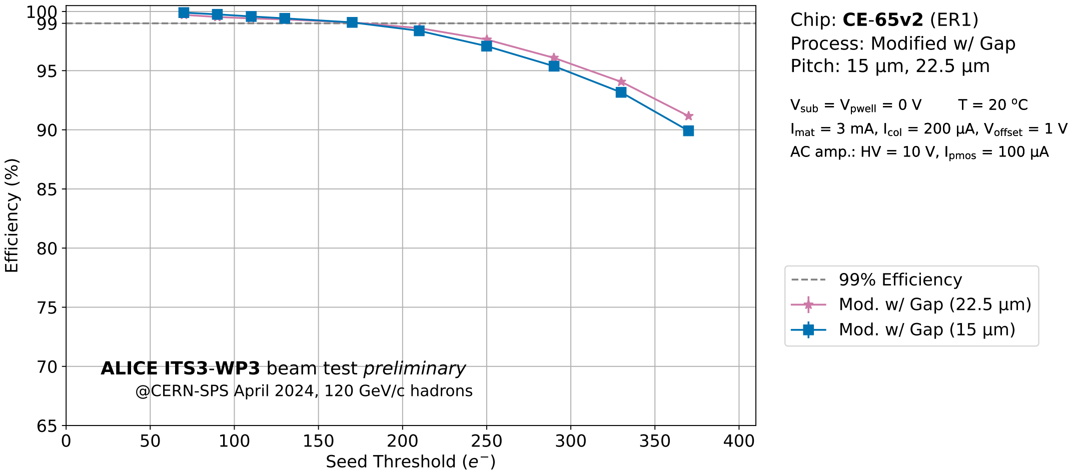


Figure 10.8: Efficiency as a function of seed threshold (e^-) for the CE-65v2 chip with a 15 µm (blue) and a 22.5 µm pitch (pink) in the Modified with Gap process. The 99% efficiency line (dotted) is marked explicitly.

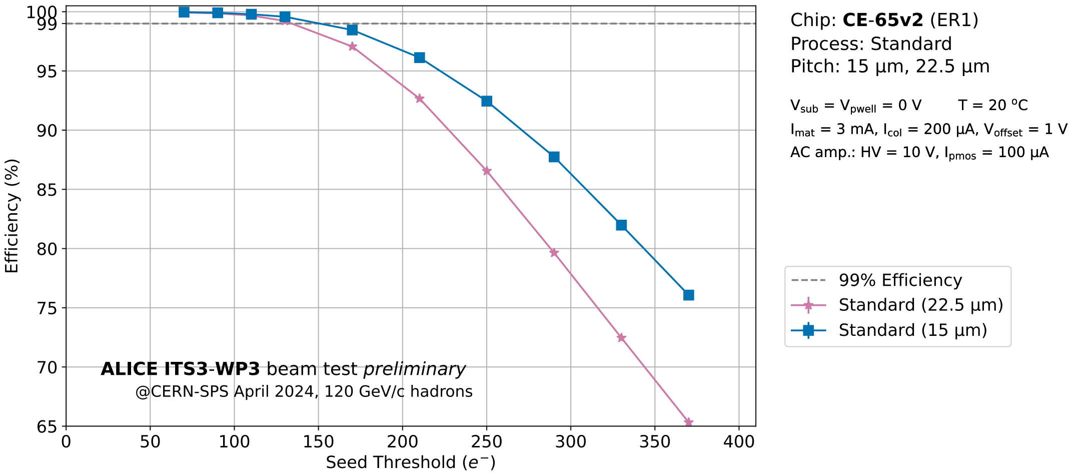


Figure 10.9: Efficiency as a function of seed threshold (e^-) for the CE-65v2 chip with a 15 µm (blue) and a 22.5 µm pitch (pink) in the Standard process. The 99% efficiency line (dotted) is marked explicitly.

degradation in efficiency, while still maintaining a high resolution throughout, as detailed earlier in this Section.

Figure 10.10 depicts the in-pixel distribution of the efficiency in the 15 µm Modified with Gap chip. The efficiency is uniform throughout the chip, remaining above 98% at all locations along the ABC paths, save for at the far corner. The uniform collection efficiency up to the pixel edges of the Modified with Gap process chips can be associated with the increased lateral electric field. The large fraction of charge that drifts towards the collection electrode makes the chip robust with respect to increasing thresholds, as shown on Figure 10.8. The 22.5 µm Modified with Gap chip shows similar trends, with the efficiency likewise rarely dropping below 98% throughout the pixel.

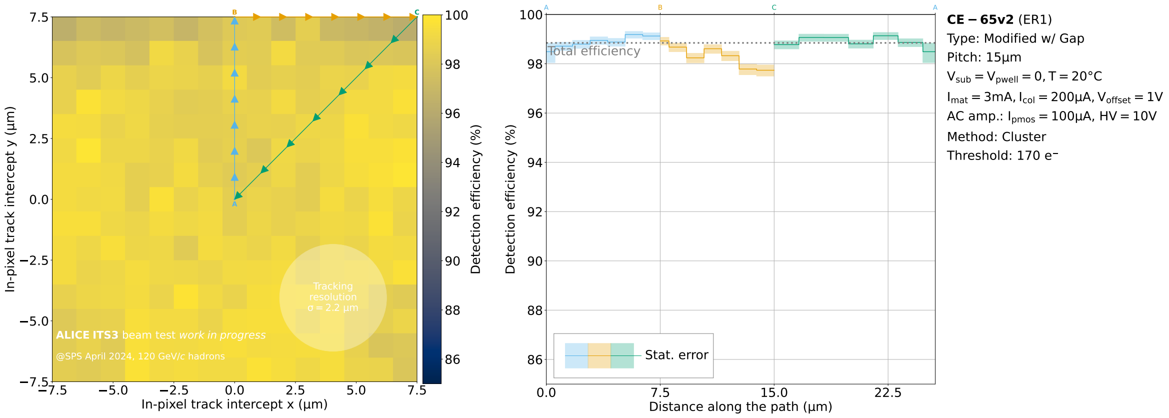


Figure 10.10: a) In-pixel efficiency distribution at a seed threshold of 170 e$^-$ for the CE-65v2 chip with a 15 µm pitch in the Modified with Gap process. b) Efficiencies extracted along three representative paths: A - towards the pixel edges, B - along the edge towards the pixel corner, and C - diagonally towards the pixel centre. The cluster method was used for reconstruction. Figure adapted from Ref. [196].

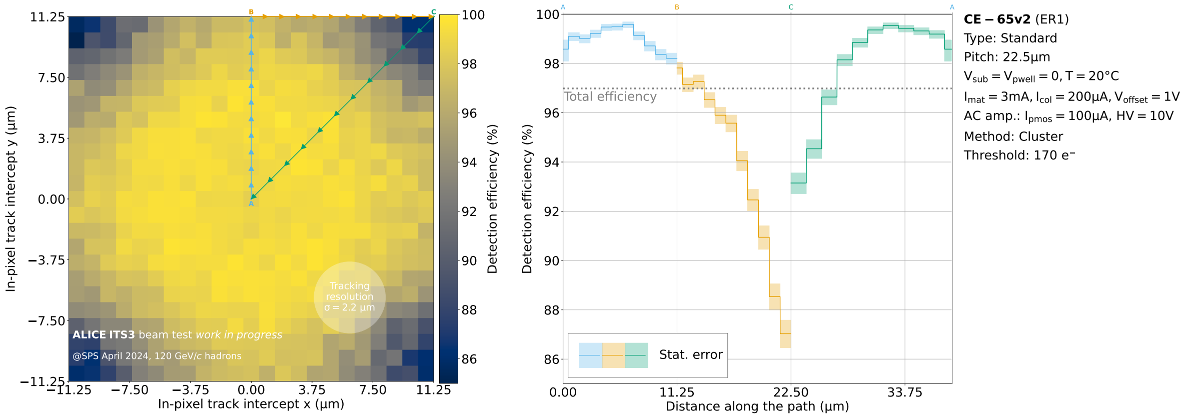


Figure 10.11: a) In-pixel efficiency distribution at a seed threshold of 170 e$^-$ for the CE-65v2 chip with a 22.5 µm pitch in the Standard process. b) Efficiencies extracted along three representative paths: A - towards the pixel edges, B - along the edge towards the pixel corner, and C - diagonally towards the pixel centre. The cluster method was used for reconstruction.

Figure 10.11 depicts the in-pixel distribution of the efficiency in the 22.5 µm Standard process chip. The efficiency distribution contrasts heavily with the Modified with Gap chips, with a clear drop towards the edges. The A path shows a gradual decrease in efficiency from ∼ 99% to ∼ 98%

as one moves towards the pixel boundary. The B path, however, accounts for the majority of the drop in efficiency to $\sim 87\%$ at the pixel corners. The diffusion-dominated charge collection at the pixel edges leads to significant charge sharing, and a corresponding drop in detection efficiency. While the same trend can be observed for the 15 µm Standard process variant, the drop in efficiency is considerably more mild to $\sim 94\%$ at the pixel edge.

10.2.4 Spatial Resolution

The spatial resolution of the CE-65v2 chip variants was studied using the window and cluster clusterisation methods. For the window method a seed threshold of ~ 100 e$^-$ was applied. For all spatial resolution results presented in this chapter, the telescope resolution is subtracted according to Equation 10.4. Figure 10.12 depicts the distribution of the x-residuals for the 4 considered chips. While the Standard process chips display true Gaussians that are well-described by the Gaussian fit, it can be seen that the Modified with Gap chips display a convolution between a Gaussian and a uniform distribution, resulting in a poor fit, particularly for the 22.5 µm chip. Nevertheless, the spatial resolution computed using the 99% sample standard deviation and the Gaussian σ varies only from 4.92 ± 0.02 µm to 4.97 ± 0.01 µm for the 22.5 µm Modified with Gap chip, and thus the Gaussian σ was used for consistency with all chips.

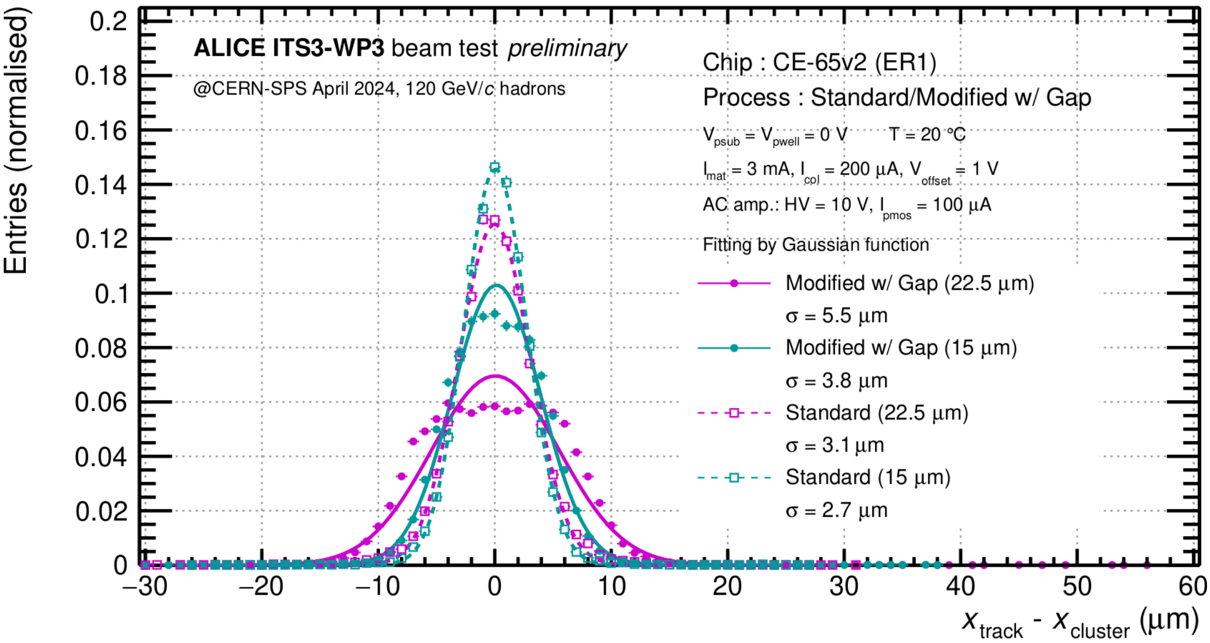


Figure 10.12: Distribution of x-residuals at a seed threshold of 100 e$^-$ for the 4 considered CE-65v2 chips. Gaussians were fit to each of the chips to extract the corresponding σ. The window method was used for reconstruction.

The Modified with Gap chips display a spatial resolution of 3.09 ± 0.02 µm and 4.97 ± 0.01 µm for the 15 µm and 22.5 µm chips, respectively. The large improvement with respect to pitch is expected due to the lack of charge sharing. In the limit of no charge sharing the improvement would be linear as expected from $\sigma = p/\sqrt{12}$ for a uniform distribution.

The Standard process chips display an excellent spatial resolution of 1.43 ± 0.02 µm and 2.02 ± 0.02 µm for the 15 µm and 22.5 µm chips, respectively. The improvement observed as a function of pitch is considerably milder in comparison to the Modified with Gap variants.

Following the hit-detection efficiency studies of Section 10.2.3, the spatial resolution was evaluated as a function of the electron threshold using the cluster method. For each threshold the analogue information passing the given threshold for the seed and its neighbours was used to construct the centre-of-mass position for the cluster. Figure 10.13 depicts the resolution as a function of electron threshold for the Modified with Gap chips. The resolution of both chips is relatively stable with increasing thresholds, increasing slightly in the 70-170 e$^-$ range, before plateauing. The 22.5 µm chip achieves a resolution slightly above 5 µm, while the 15 µm achieves a resolution in the $\sim$ 3.5 µm range. For both chips a resolution considerably better than the binary resolution of $p/\sqrt{12}$ is achieved.

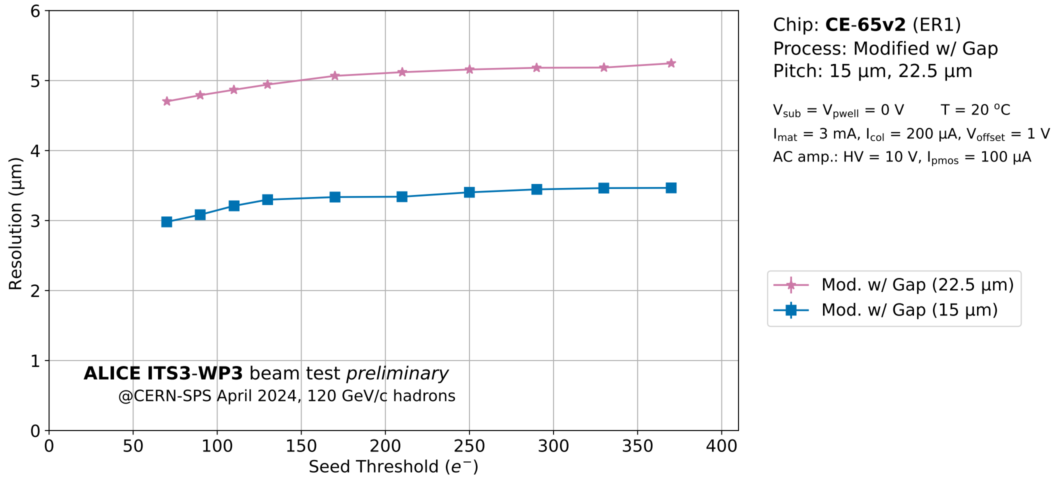


Figure 10.13: Resolution as a function of seed threshold (e$^-$) for the CE-65v2 chip with a 15 µm (blue) and a 22.5 µm pitch (pink) in the Modified with Gap process.

Figure 10.14 depicts the resolution as a function of electron threshold for the Standard process chips. The resolution of both chips shows a sizeable improvement with respect to their Modified with Gap counterparts. At a seed threshold of 70 e$^-$ an excellent resolution of $\sim$ 1.4 µm and $\sim$ 1.8 µm is achieved. The resolution of both pitch chips degrades quickly in the 70-250 e$^-$ range as the cluster size decreases, before plateauing at slightly above 5 µm and $\sim$ 3.5 µm for the 22.5 µm and 15 µm pitch chips, in parity with the Modified with Gap process chips.

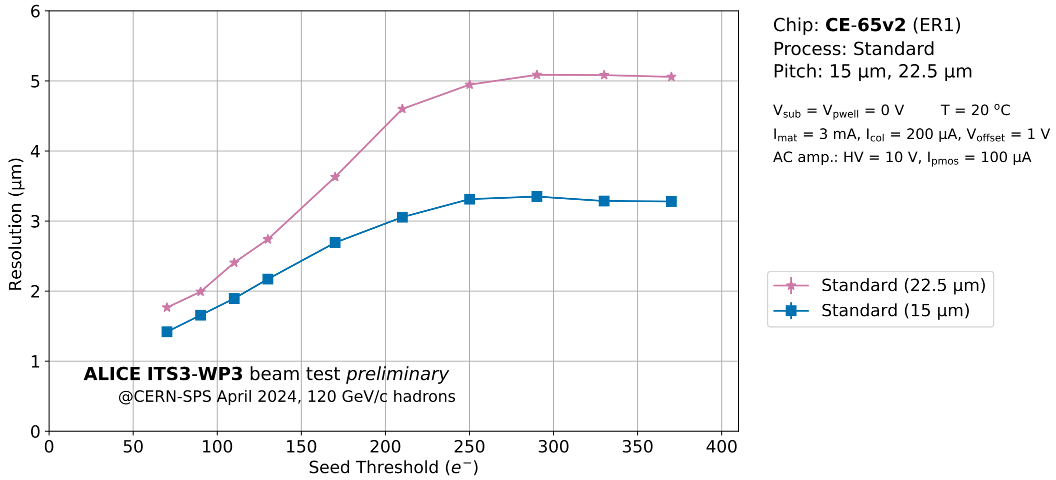


Figure 10.14: Resolution as a function of seed threshold (e$^-$) for the CE-65v2 chip with a 15 µm (blue) and a 22.5 µm pitch (pink) in the Standard process.

Importance of Digitisation

The precise determination of the measured track position requires the full exploitation of the sensor's signal. In the CE-65 family of chips, this entails the readout of the full analog signal, which, while referred to as "analog" throughout this thesis, is not "analog" in the formal sense, as it is digitised[2] to high precision by the 16-bit ADC introduced in Section 9.1.1. The readout of the "analog" signal presents a data rate challenge. Typically this is addressed by setting a threshold (much like the electron thresholds explored in Section 10.2.3 and earlier in this Section), above which a binary or time over threshold signal corresponding to a hit is read out.

In this section the importance of digitising the signal was explored by performing the cluster reconstruction using a variety of methods. Firstly, a lower bound on the performance of a clusterisation method was set by considering the spatial resolution achievable if only the seed position is used for reconstruction. This is referred to as the *seed* method. Next, in the *binary* method, a cluster was built using the *cluster* method. However, instead of the analogue information, the hit or no-hit information was used for computing the centre of the cluster.

The *cluster* and *window* clusterisation methods were used to set an upper bound on the achievable spatial resolution. The ADC range from [0, 10000] was digitised to $n = 1, 2, 4, 6, 8, 16$ bits for the window method in order to observe the degradation in the achievable spatial resolution when a coarser reconstruction is performed. Digitisation was emulated during analysis by performing a trivial bit-shift, and considering only the n leading bits. The ADC range from [threshold, 10000] was similarly digitised to $n = 0, 1, 2, 4, 6, 8, 16$ bits for the cluster method in order to perform the same comparison. Where relevant for each clusterisation procedure, the seed and neighbour thresholds were set to 100 e^-.

Figure 10.15 depicts the achievable resolution for each of the clusterisation methods for the 15 µm Modified with Gap and 22.5 µm Standard process chips.

For the 15 µm Modified with Gap process chip a resolution of just above 3 µm is achieved with the binary method. Such a low resolution can be associated with tracks close to the pixel boundaries where considerable charge is shared with the neighbours, even in the Modified with Gap process, aiding in the cluster reconstruction. Remarkably, additional information during clusterisation has a marginal improvement in the precision of the reconstruction. As more bits are included in the cluster method, and even in the analog limit, the achievable resolution hardly changes. The window method performs poorly for $n < 4$ bits. Presumably, the most important information is encoded at low ADC counts given that low amounts of charge are generated by MIPs traversing the epitaxial layer, as shown in Figure 10.6. This information is neglected during digitisation, due to the linear spread in digitisation bins. For $n = 6, 8, 16$ bits, the window method outperforms the analog window method. This was observed for both the 15 µm and 22.5 µm Modified with Gap chips. While the $n = 16$ bit window method might be expected to be equivalent to the analog window method, the digitised window method does not include negative signals resulting from the CDS, which are set to 0 during the digitisation procedure. This results in a reduction in noise during the cluster reconstruction.

The 22.5 µm Standard process chip shows a much larger improvement in the spatial resolution when moving from the seed clusterisation method to the binary method, in accordance with the

---

[2] "Digitisation" is used throughout this section to refer exclusively to the signal quantisation step, and should not be confused with the broader signal processing chain.

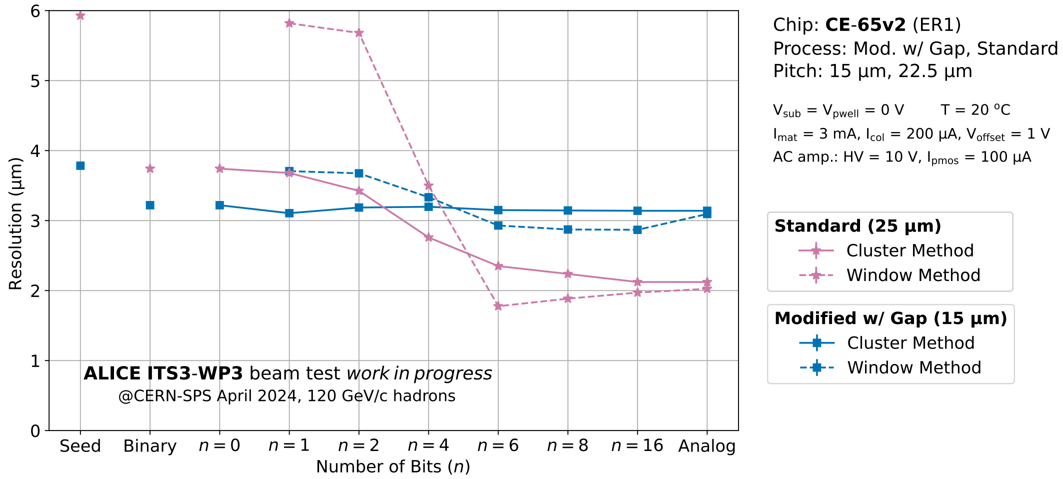


Figure 10.15: Resolution as a function of clusterisation method at a seed threshold of 100 e$^-$ for the CE-65v2 chip with a 15 µm pitch in the Modified with Gap process (blue) and a 22.5 µm pitch in the Standard process (pink). Variants of the window method with different digitisation granularities are connected by dashed lines. Moving from left to right corresponds to an increased number of digitiation bits, until the full analog resolution is reached. The corresponding resolutions for the cluster method are connected by the solid lines.

high charge sharing evident in the Standard process. The effect is particularly pronounced until $n = 6$ bits, where a spatial resolution of 2.35 µm is achieved. This contrasts heavily with the Modified with Gap process chips, where additional bits lead to a negligible benefit. The window method, by comparison, initially shows poor performance for $n = 1, 2$, before rapidly improving for $n = 4, 6$. At $n = 6$ a resolution 1.78 µm is achieved. For $n > 6$ no further gain in the spatial resolution is observed, which begins to degrade slightly, potentially due to unfortunate bin boundaries for these n. The 15 µm Standard process chip likewise shows this rapid improvement in the $n = 2 - 6$ bit range, but does not show the slight degradation above $n = 6$.

In-pixel Spatial Resolution

Similarly to Section 10.2.3, a threshold of ~ 170 e$^-$ was chosen in order to study the in-pixel distribution of the mean absolute deviation, as a proxy for the spatial resolution. In order to maximally exploit the capabilities of the chip, the window method was used for clusterisation.

The mean absolute deviation Δr is defined by summing the residuals in the x and y dimensions in quadrature:

$$\Delta r = \sqrt{(\Delta x)^2 + (\Delta y)^2} \tag{10.10}$$

and subtracting the telescope resolution, as in Equation 10.4. Assuming Gaussian distributions for the residuals of x and y with equal standard deviations $\sigma_{\Delta x} = \sigma_{\Delta y} := \sigma$, and zero mean, as in Figure 10.12 for all variants save for the 22.5 µm Modified with Gap chip, the distribution of Δr is given by the Rayleigh density [222]. The expectation value of the Rayleigh density has the well-known value of $\mathbb{E}[\Delta r] = \sqrt{\frac{\pi}{2}} \cdot \sigma$. Thus, the mean absolute deviation can be interpreted as a scaled estimate of the spatial resolution.

Figure 10.16 depicts the in-pixel distribution of the mean absolute deviation in the 15 µm Modified with Gap chip. The mean absolute deviation achieves its lowest values in the central bins, where tracks are likely to deposit their full charge in a single pixel, leading to a trivial reconstruction of the position. Moving away from the centre of the pixel along the A path leads to a rapid deterioration in the achievable mean absolute deviation to ∼ 5 µm at the boundary of the pixel. The B path, towards the edge of the pixel, leads to a further decline to ∼ 6 µm. The large variance in the achievable mean absolute deviation suggests that only central tracks are identified with a high precision. The fast deterioration of the mean absolute deviation is also observed for the 22.5 µm chip, where a mean absolute deviation of above ∼ 6 µm is reached towards the corners of the chip. Hits towards the corner of the chips lie the furthest from the collection electrode, leading to the larger residuals for single-pixel events, common in the Modified with Gap process.

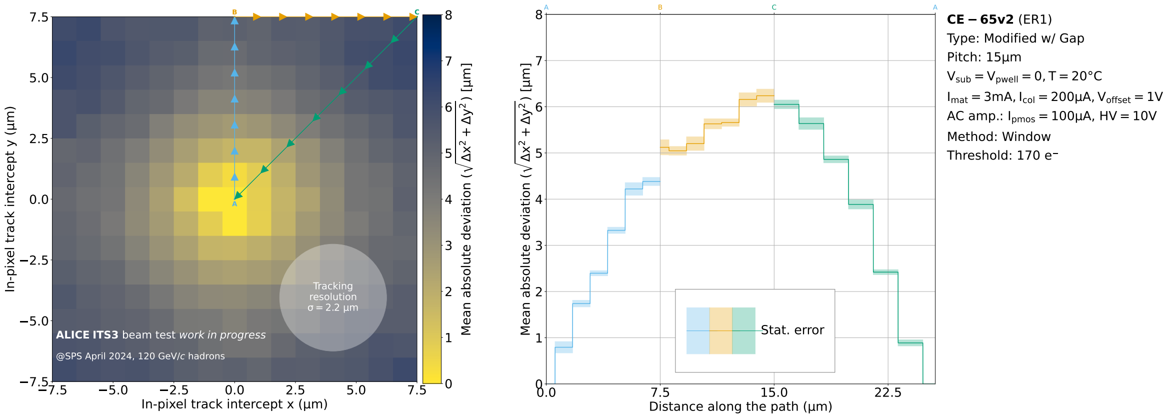


Figure 10.16: a) In-pixel mean absolute deviation distribution at a seed threshold of 170 e⁻ for the CE-65v2 chip with a 15 µm pitch in the Modified with Gap process. b) Mean absolute deviation extracted along three representative paths: A - towards the pixel edges, B - along the edge towards the pixel corner, and C - diagonally towards the pixel centre. The window method was used for reconstruction. Figure adapted from Ref. [196].

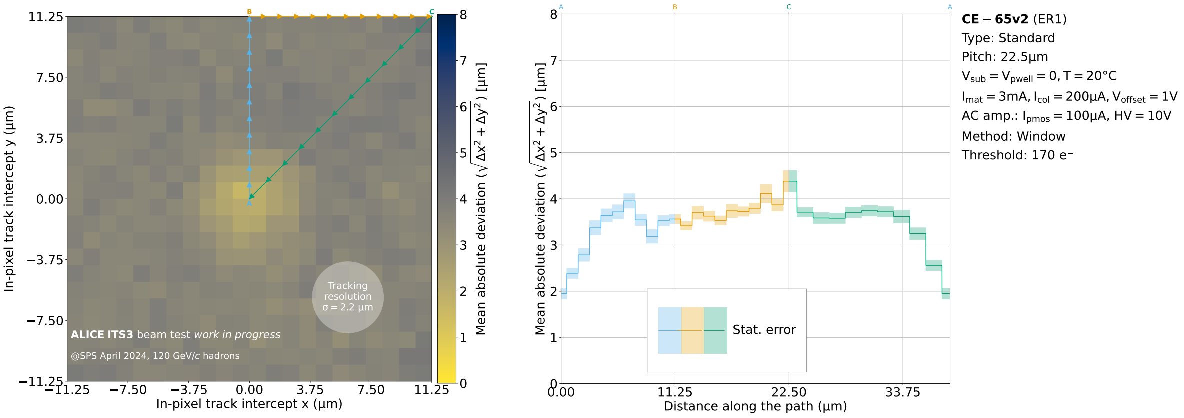


Figure 10.17: a) In-pixel mean absolute deviation distribution at a seed threshold of 170 e⁻ for the CE-65v2 chip with a 22.5 µm pitch in the Standard process. b) Mean absolute deviation extracted along three representative paths: A - towards the pixel edges, B - along the edge towards the pixel corner, and C - diagonally towards the pixel centre. The window method was used for reconstruction.

Figure 10.17 depicts the in-pixel distribution of the mean absolute deviation in the 22.5 µm Standard process chip. The mean absolute deviation is considerably more uniform than for the

Modified with Gap process chips. In particular, the mean absolute deviation begins to degrade when moving along the A path away from the centre to ~ 3.5 µm. Beyond $\sim p/4$ from the centre, the mean absolute deviation hardly varies, including at the pixel edges, suggesting an accurate position reconstruction for high charge sharing events. The same trend can be observed for the 15 µm Standard process variant, with the mean absolute deviation stabilizing around ~ 3 µm outside the pixel centre, and decreasing below the telescope resolution therein.

10.2.5 Charge Sharing

Charge sharing behaviour significantly impacts the hit-detection efficiency and the spatial resolution, making it a key metric when comparing different pixel configurations. To study the extent of charge sharing for the different CE-65v2 variants, the accumulated charge ratio, introduced in Section 10.1.1, was used. It is defined as the fraction of charge carried by the n first pixels in a cluster.

Figure 10.18a and 10.18b depict the accumulated charge ratio as a function of the number of pixels in the cluster for the 15 µm and 22.5 µm pitch Modified with Gap process CE-65v2 chips. The average charge fraction collected by the central pixel of the two chips is 83.9% and 86.8%, respectively, demonstrating the low charge sharing characteristic of the Modified with Gap process. When considering the two most energetic pixels, both the average and most probable value of the accumulated charge are around unity for both chips, indicating that the vast majority of events consist of single-pixel or two-pixel clusters. Indeed, if a hit occurs close to the pixel boundary, then significant charge may diffuse a neighbouring pixel, otherwise the charge is almost entirely collected in the central pixel due to the strong drift current of the Modified with Gap process. Due to the Gap modification at the pixel boundaries, the electric field propagates well. Coupled with a larger pitch, the 22.5 µm pitch Modified with Gap chip exhibits the lowest charge sharing of all considered chip variants.

Figure 10.19a and 10.19b show the accumulated charge ratio for the 15 µm and 22.5 µm Standard process chips. The average charge fraction collected by the central pixel in the 15 µm Standard process chip is 66.1%, indicating substantial charge sharing. The 22.5 µm pitch chip exhibits the most extreme charge sharing of all considered variants, with an average charge fraction of less than 60%, and the most probable value being even lower at $\sim 45\%$. This results from the competing effects of a larger pitch, whereby more charge is collected due to the larger area, and the electric field not propagating well at the pixel edges, resulting in considerably less charge drift. It should be noted that due to charge ordering, negative noise contributions resulting from the CDS can appear in the larger cluster size bins towards the right side of Figures 10.18 and 10.19. These negative noise contributions can result in the accumulated charge exceeding 100%.

Considering the two *exploration axes* of process modification and pixel pitch, the SPS testbeam quantitatively substantiated that the Standard process results in considerably more charge sharing than its Modified with Gap counterpart. The dependence on the pixel pitch is inverted for the processes, with a larger pitch resulting in less charge sharing for the Modified with Gap process, but a larger pitch resulting in more charge sharing for the Standard process. The effect of biasing on charge sharing was explored at the DESY testbeam, detailed in Section 10.3.2.

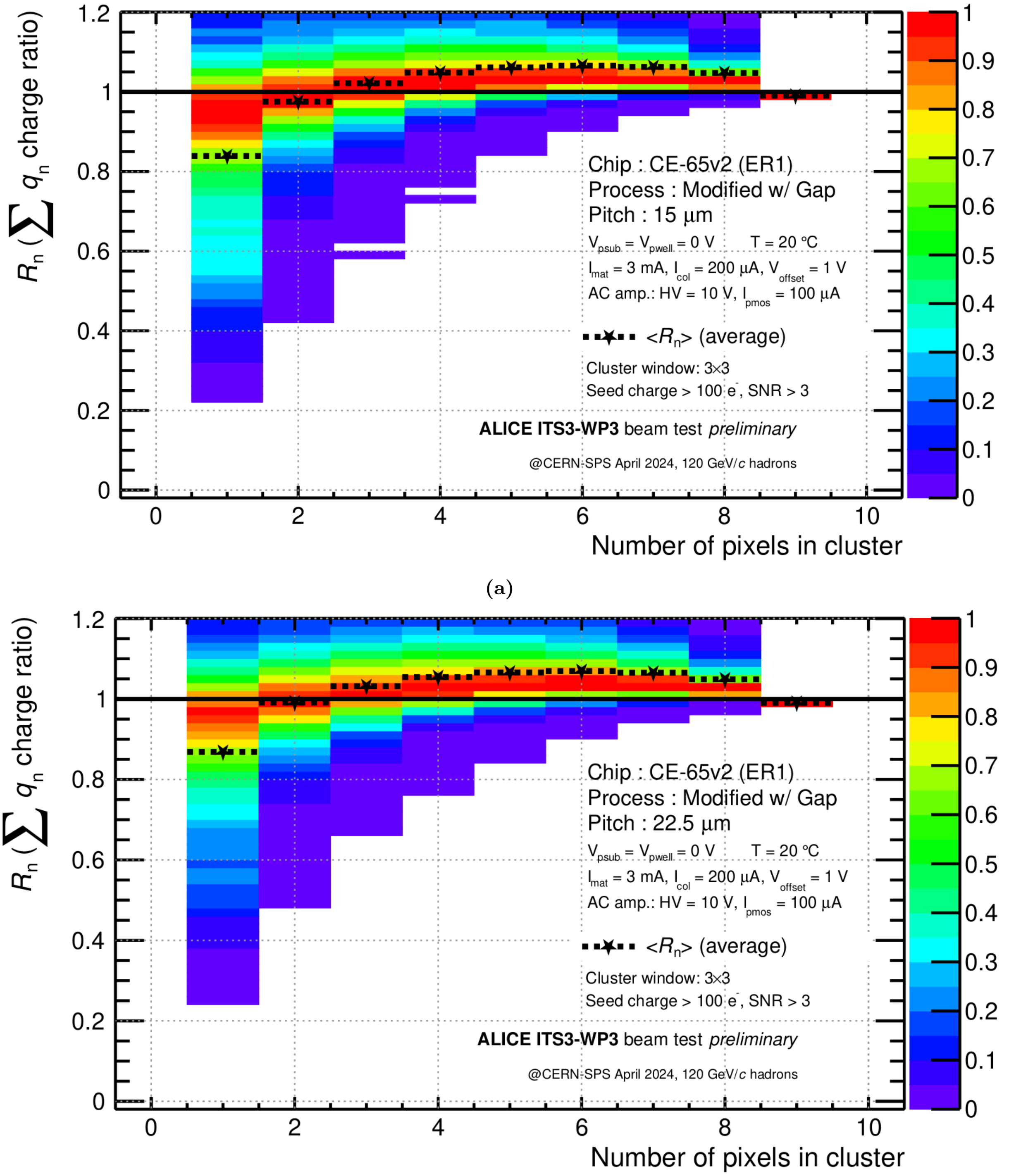
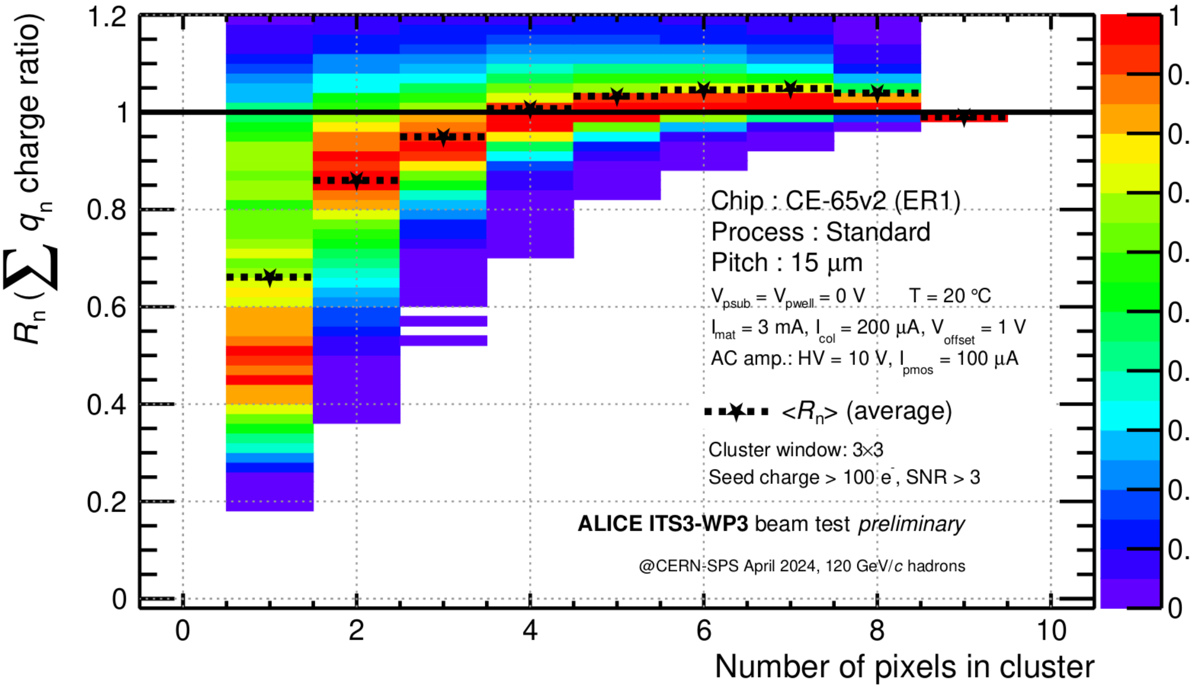
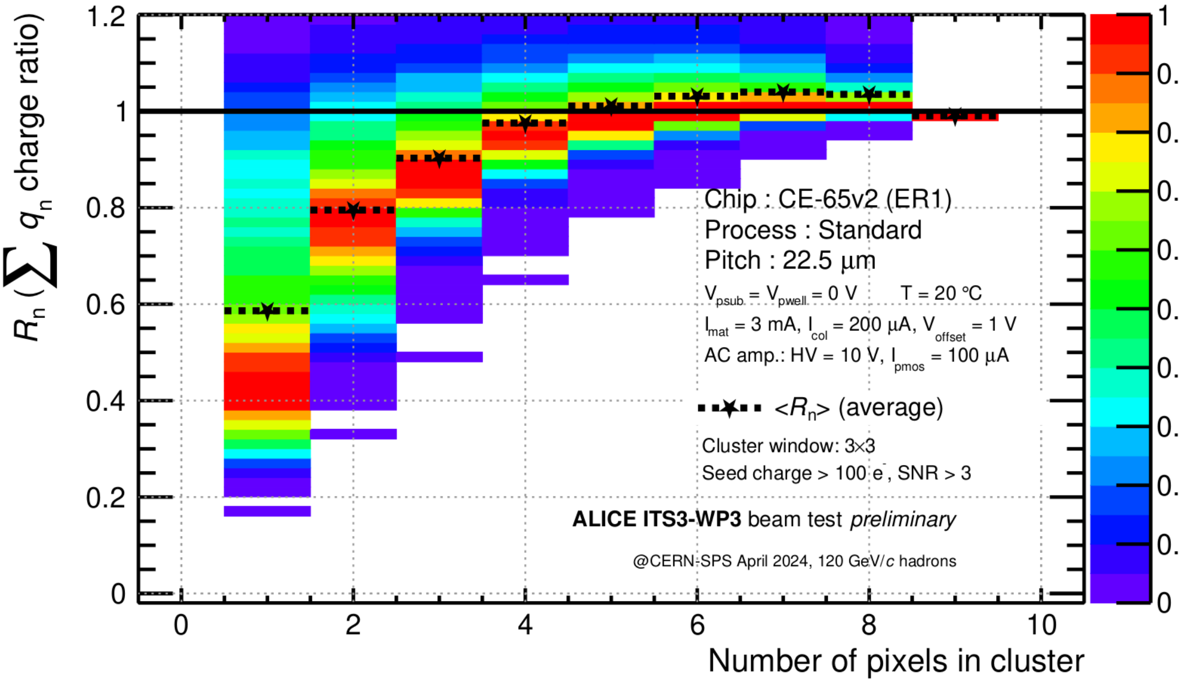


Figure 10.18: Accumulated charge ratio as a function of number of pixels in a cluster for the CE-65v2 chip with a 15 µm pitch (a) and a 22.5 µm pitch (b) in the Modified with Gap process.



(a)



(b)

Figure 10.19: Accumulated charge ratio as a function of number of pixels in a cluster for the CE-65v2 chip with a 15 µm pitch (a) and a 22.5 µm pitch (b) in the Standard process.

10.2.6 Ramifications of a Staggered Matrix Geometry

The staggered matrix geometry reduces the number of immediate neighbours from 8 to 6. Nevertheless, in the staggered matrix geometry, the detector channels no longer define a Cartesian coordinate system. In order to perform the vector addition necessary for determining the cluster centre, the detector channels were first transformed to the *local* Cartesian coordinate system, introduced in Section 10.1.2, as depicted on Figure 10.3. The analysis strategy employed for the staggered matrix geometry chips mirrored the strategy employed for the square chips. Both the window and the cluster methods were used for the determination of the spatial resolution and, for the latter, efficiency. For the window method, the cluster centre was defined as the centre-of-mass position of the seed pixel and its 6 immediate neighbours. For the cluster method, the size-n clusterisation procedure detailed in Section 9.3.2 was used, with the immediate neighbours also defined by the 6 pixels around the seed pixel.

The window clusterisation method with a seed threshold[3] of $\sim$ 100 e^- was used for the analysis of 22.5 µm and 18 µm Modified with Gap and Standard process staggered matrix chips. Table 10.1 summarizes the obtained spatial resolution for the eight chips analysed in this thesis.

Table 10.1: Measured spatial resolution for the squared and staggered CE-65v2 variants at the SPS testbeam at a seed threshold of 100 e^- using the window clusterisation method. The improvement of the staggered variants is mainly driven by the y-residuals. Uncertainties reflect the statistical error from the fit.

Matrix	Process	Pitch (µm)	Spatial Resolution (µm)
Square	Mod. w/ Gap	22.5	4.97 ± 0.1
	Mod. w/ Gap	15	3.09 ± 0.2
	Standard	22.5	2.02 ± 0.2
	Standard	15	1.43 ± 0.2
Staggered	Mod. w/ Gap	22.5	4.62 ± 0.1
	Mod. w/ Gap	18	3.50 ± 0.2
	Standard	22.5	1.94 ± 0.2
	Standard	18	1.49 ± 0.2

Comparing the 22.5 µm staggered matrix chips to their square matrix counterparts reveals an improvement in the spatial resolution. The improvement is pronounced for the Modified with Gap chip, which improves from 4.97 ± 0.1 µm to 4.62 ± 0.1 µm.

Figure 10.20 depicts the x-residuals for each of the four staggered matrix chips. Similar trends to the square matrix chips can be observed, with the Standard process chips achieving a significantly better resolution than the Modified with Gap variants. Smaller pitches lead to an improved resolution, with the 18 µm Standard process chip now achieving the best x-residuals at 1.49 ± 0.2 µm.

Figure 10.21 depicts the y-residuals for each of the staggered matrix chips. The residual distributions for all four chips are narrow. For the Standard process chips this entails a decrease of 0.3 µm for both the 22.5 µm and 18 µm chips, with respect to their x-residual distribution. For the Modified with Gap chips the decrease is of 0.4 µm and entails a change in the shape of the distribution. While the x-residuals of the Modified with Gap process chips retain their

---

[3] As the gain calibration for the staggered chips was not completed at the time of writing, the γ_{ccf} values from the square chips were used. Given that γ_{ccf} does not depend on matrix geometry, the resulting error is expected to be negligible.

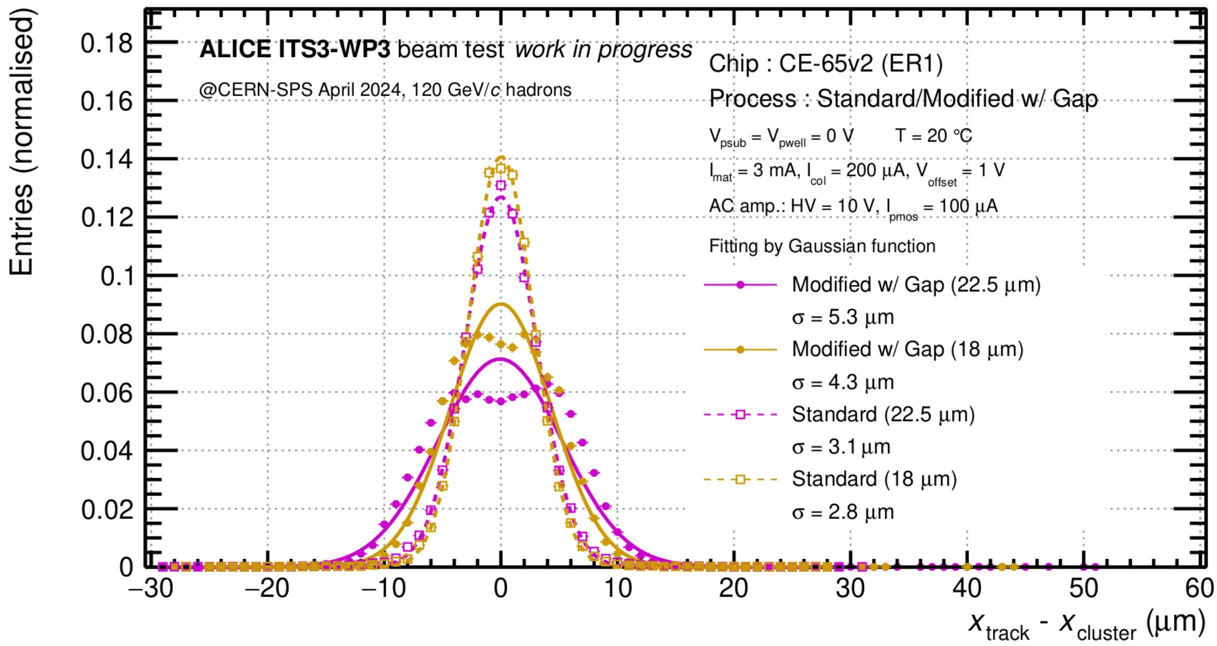


Figure 10.20: Distribution of x-residuals at a seed threshold of 100 e$^-$ for the 4 considered staggered matrix CE-65v2 chips. Gaussians were fit to each of the chips to extract the corresponding σ. The window method was used for reconstruction.

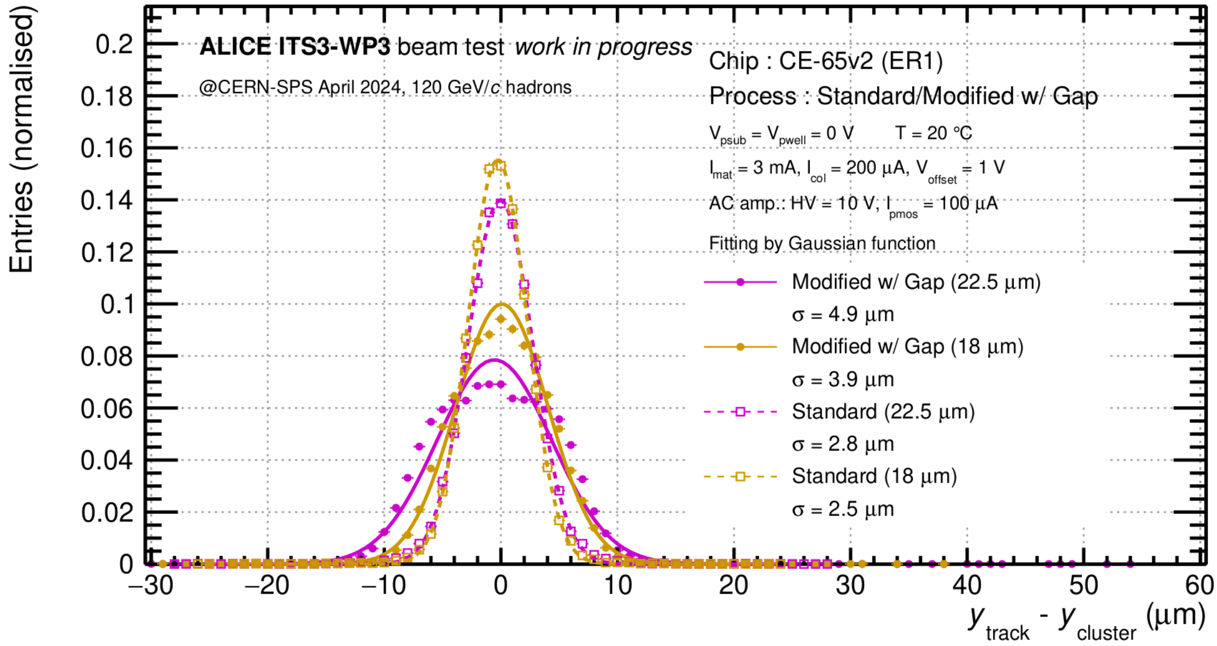


Figure 10.21: Distribution of y-residuals at a seed threshold of 100 e$^-$ for the 4 considered staggered matrix CE-65v2 chips. Gaussians were fit to each of the chips to extract the corresponding σ. The window method was used for reconstruction.

characteristic shape of the convolution between a Gaussian and a step function, the y-residuals have higher central values about the origin, returning to a slightly more Gaussian shape.

Hit-Detection Efficiency

Electron threshold scans were performed for all four of the staggered matrix chips in order to study the ramifications of the staggered matrix geometry on the development of both the hit-detection efficiency and the spatial resolution. As in Section 10.2.4, the cluster method was used with the seed and neighbour thresholds being equal.

Figure 10.22 depicts the efficiency as a function of threshold for the 22.5 µm and 18 µm Modified with Gap staggered matrix chips. In addition, the 22.5 µm Modified with Gap square matrix chip is included for comparison. Comparing the square and staggered matrix 22.5 µm chips reveals a marginal increase in efficiency towards higher thresholds. In accordance with the trends observed in Section 10.2.3 for the square matrix chips, the 18 µm staggered matrix chip shows a slightly lower efficiency than its larger pitch counterpart.

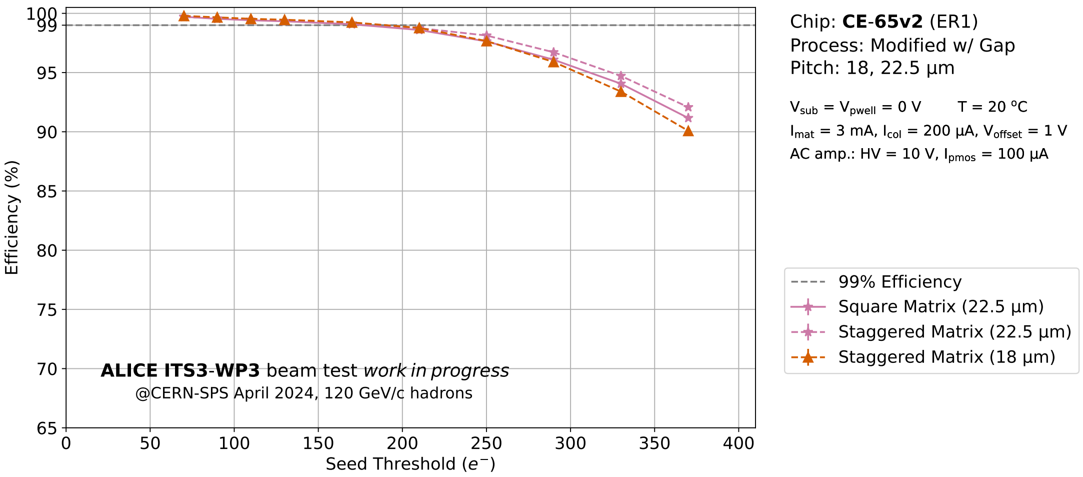


Figure 10.22: Efficiency as a function of seed threshold (e^-) for the staggered matrix CE-65v2 chip with a 18 µm (orange) and a 22.5 µm pitch (pink) in the Modified with Gap process. For reference, the 22.5 µm squared matrix arrangement CE-65v2 chip is included as a solid line. The 99% efficiency line (dotted) is marked explicitly.

A similar trend can be observed for the Standard process staggered matrix chips, as depicted on Figure 10.23. The 22.5 µm staggered matrix chip once again shows a slight increase in efficiency. The increase can be observed even at low thresholds, offering a marginally increased operating range. The 18 µm pitch chip, with its correspondingly lower charge sharing, retains a considerably higher efficiency at higher thresholds.

Spatial Resolution

The spatial resolution as a function of electron threshold is depicted for the 22.5 µm and 18 µm Modified with Gap staggered matrix chips, and the 22.5 µm square matrix counterpart. As for previous threshold scan sections, the cluster method is used for reconstruction. The staggered matrix 22.5 µm chip shows virtually no gain with respect to the square arrangement. At low thresholds, a mild improvement in the y-residuals manifests as a slightly better spatial

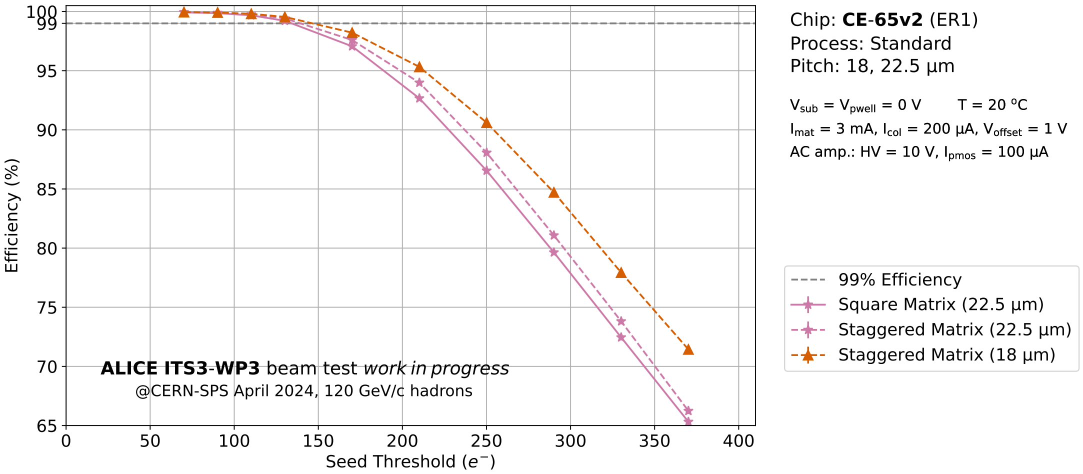


Figure 10.23: Efficiency as a function of seed threshold (e^-) for the staggered matrix CE-65v2 chip with a 18 µm (orange) and a 22.5 µm pitch (pink) in the Standard process. For reference, the 22.5 µm squared matrix arrangement CE-65v2 chip is included as a solid line. The 99% efficiency line (dotted) is marked explicitly.

resolution, but this effect is not as pronounced as for the window method, summarized on Table 10.1. The 18 µm staggered matrix chip achieves a fairly stable resolution of ∼ 4 µm, which would place it between the resolutions achieved by the 15 µm and 22.5 µm Modified with Gap process square matrix chips, depicted on Figure 10.24.

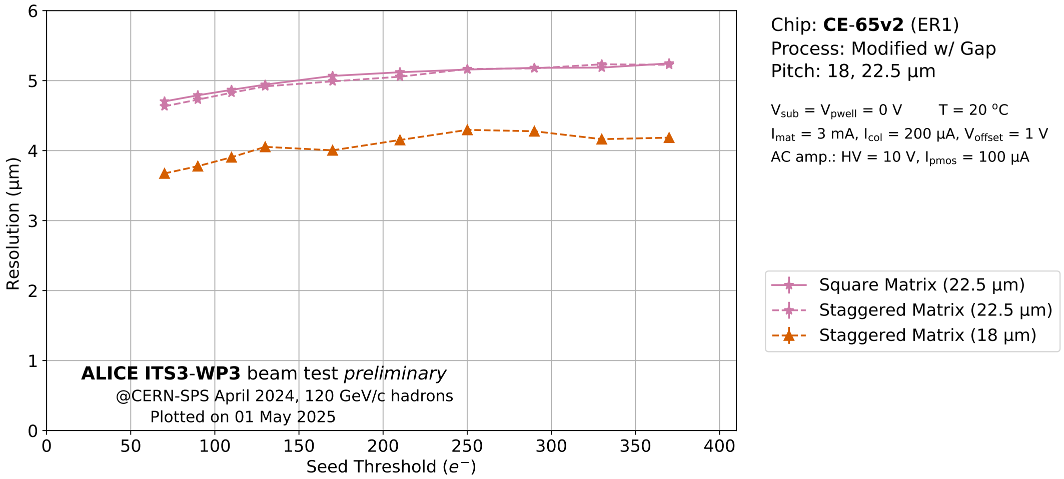


Figure 10.24: Resolution as a function of seed threshold (e^-) for the staggered matrix CE-65v2 chip with a 18 µm (orange) and a 22.5 µm pitch (pink) in the Modified with Gap process. For reference, the 22.5 µm squared matrix arrangement CE-65v2 chip is included as a solid line.

The spatial resolution obtained by the Standard process chips benefits sizeably more from the staggered arrangement, as depicted on Figure 10.25. In particular, the 22.5 µm staggered matrix chips achieves a lower spatial resolution than its square counterpart. While the obtained spatial resolution converges for very low thresholds of < 100 e^- or > 300 e^- for both matrix arrangements, the staggered matrix achieves a considerably lower spatial resolution between the extremes. The measured improvement is maximal at a threshold of 210 e^-, where the staggered matrix arrangement achieves a spatial resolution that is 16% better than the square arrangement.

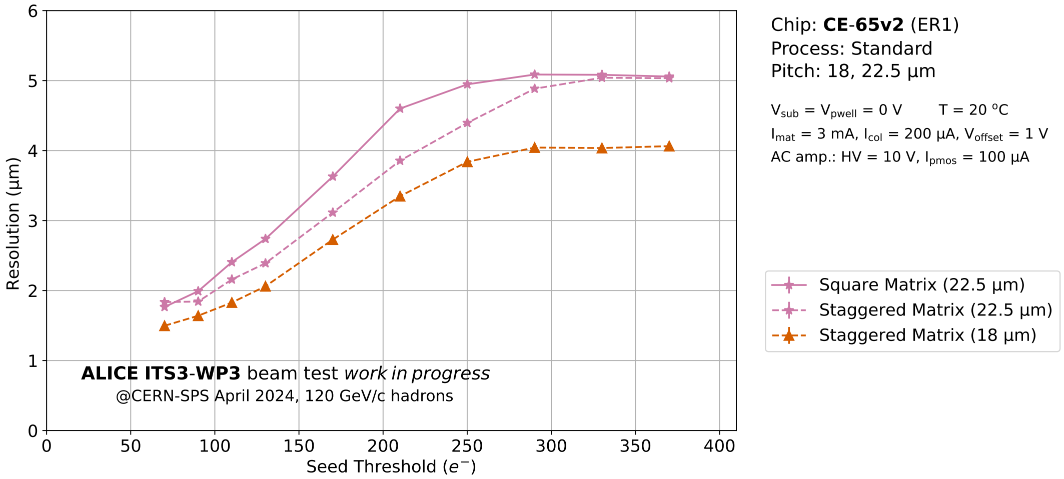


Figure 10.25: Resolution as a function of seed threshold (e^{-}) for the staggered matrix CE-65v2 chip with a 18 µm (orange) and a 22.5 µm pitch (pink) in the Standard process. For reference, the 22.5 µm squared matrix arrangement CE-65v2 chip is included as a solid line.

Effect of Matrix Geometry on Cluster Size and Reconstruction Error

The superior resolution observed for the staggered matrix layout can be attributed to both reduced charge sharing among neighbours and the inherent geometric advantages of the arrangement. In particular, the seed pixel of a given cluster for the square and staggered matrix arrangements was found to carry a similar fraction of the cluster signal at $\sim 60\%$ for the Standard process. However, the remaining fraction of the cluster signal is spread among only 6 neighbours in the staggered arrangement, as opposed to the 8 of the square arrangement. This results in an improved position reconstruction owing to the typical cluster size distribution. Figure 10.26 depicts the cluster sizes of the Standard process chips at the lowest measured threshold in the cluster method of 70 e^{-} for both the square and staggered arrangements. In the square arrangement the most common cluster sizes are $n = 4$ and $n = 2$, as observed for initial studies of the ALPIDE chip [212]. Geometrically, the $n = 2$ cluster corresponds to the seed pixel and one of its immediately adjacent neighbours, which does not necessarily entail a poor position reconstruction if the hit is towards the centre of the boundary of the seed pixel and its neighbour. The $n = 4$ cluster distribution is dominated by the "square" cluster whereby a 2×2 window containing the seed defines the cluster [223].This is associated with hits toward the seed edges, where the charge diffuses to the three neighbours. While this enables a precise position reconstruction, as observed by the excellent resolution the square matrix arrangement achieves at low thresholds, it is not robust with respect to increasing thresholds, as the shared charge is diluted among three pixels.

This can be observed in Figure 10.26 where at 210 e^{-} the size-4 clusters already form a sizeably smaller fraction of the cluster size distribution. The 22.5 µm Standard process staggered arrangement chip, on the other hand, contains almost double the fraction of size-3 cluster events as its square arrangement counterpart at a threshold of 70 e^{-}. Indeed, the size-3 clusters now constitute the most common type of cluster. While this entails a smaller cluster, and thus potentially reduced reconstruction precision, than the size-4 clusters which commonly appear for the square matrix at low thresholds, the 2 neighbours of the size-3 clusters of the staggered arrangement will contain more charge than the 3 neighbours of the size-4 clusters, and thus

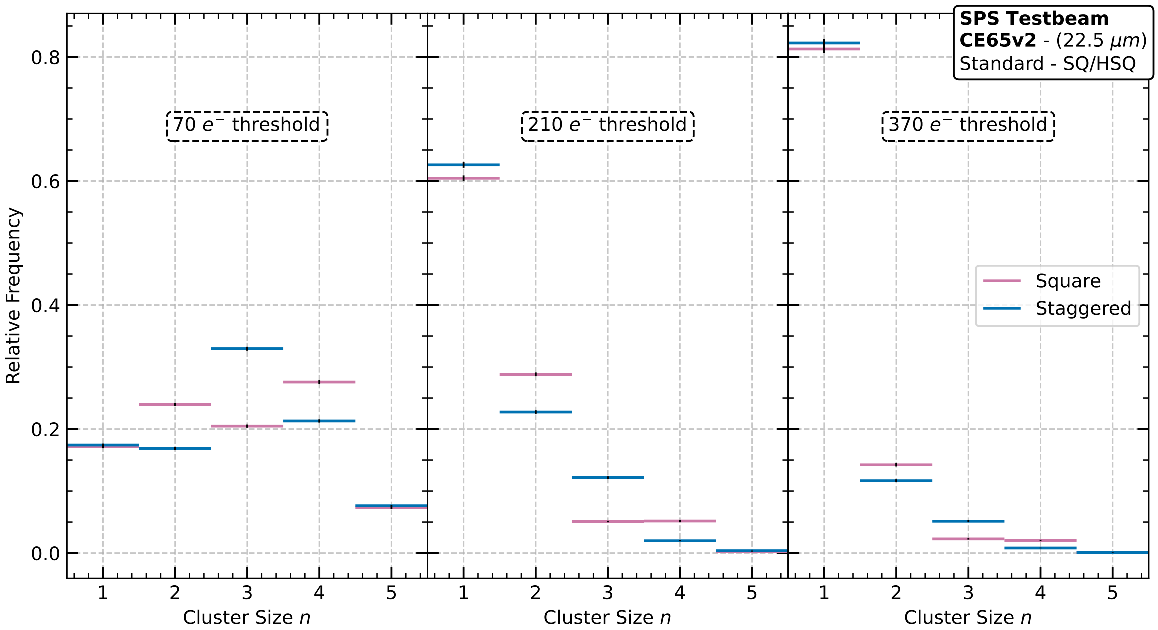


Figure 10.26: Cluster size distributions at seed thresholds of 70 e$^-$ (a), 210 e$^-$ (b), 370 e$^-$ (c), for the square and staggered matrix CE-65v2 chips with a 22.5 µm pitch in the Standard process.

be more robust towards higher thresholds. This is observed in the cluster size distribution at a threshold of 210 e$^-$, depicted on Figure 10.26. The size-3 clusters constitute 12.2% of the threshold for the staggered arrangement, whereas for the square chip the size-3 clusters constitute only 5.1% of the cluster distribution. Accounting for clusters of size-n>3 increases this figure to 10.7% for the square arrangement, which is still below the observed fraction of size-3 clusters for the staggered arrangement.

The increase in n>3 size clusters does not account for the improvement in resolution alone. A geometric argument for the observed improvement in the resolution of the staggered matrix can be made even for size-2 clusters. Considering still the cluster size distribution at a threshold of 210 e$^-$, it can be seen that after single-pixel clusters, clusters of size-2 are the most common. In the square arrangement this corresponds to clusters consisting of the seed pixel and its immediately adjacent neighbour in the $\pm x$ or $\pm y$ directions, where the y dimension is slightly more common [223]. The size-2 clusters extending in the y dimension in the staggered arrangement are identical, as the staggering is performed in the vertical direction, and thus no difference is expected for these. However, the remaining size-2 clusters consist of the seed pixel, and one of the 2 neighbours in either the $+x$ or $-x$ direction. Without loss of generality, consider a hit at (x^*, y^*) towards the right boundary of upper-half of a seed pixel, as depicted by the red cross on the left of Figure 10.3. The resulting size-2 cluster will consist of the seed pixel at (x, y) and its neighbour at ($x + p$, $y + p/2$). Projecting the measured positions of the cluster on the x-axis results in two distinct positions. The same is true for the projection of the cluster positions on the y-axis. In the square arrangement, however, the projection of the cluster positions on the y-axis reveals that there there is only one position, resulting in a poorer reconstruction of the y position.

This argument can be formalized somewhat by considering the expected error in the y dimension for each of the arrangements. It follows trivially that the reconstructed position in a size-2 cluster where a fraction A of the cluster charge is shared, is given by

$$(x_{cl.}, y_{cl.}) = A \cdot (x_2, y_2) + (1 - A) \cdot (x_1, y_1) = (x_1, y_1) + A \cdot (x_2 - x_1, y_2 - y_1) \tag{10.11}$$

and thus lies a fraction A along the diagonal connecting the two pixels. In the limit of perfect charge sharing the reconstructed position is given by the projection of the hit position (x^*, y^*) onto the linking vector $\vec{v}$.

For the staggered matrix the linking vector v becomes

$$\vec{v} = (x_2 - x_1, y_2 - y_1) = (p, p/2) \tag{10.12}$$

Setting the origin of the coordinate system at (x_1, y_1) means the reconstructed position is given by

$$\begin{aligned}(x_{cl.}, y_{cl.}) &= \frac{(x^*, y^*) \cdot (p, p/2)}{|(p, p/2)|^2}(p, p/2) = \frac{x^*p + y^*p/2}{p^2 + p^2/4}(p, p/2) = \frac{4x^* + 2y^*}{5}(1, 1/2) \\ &= \left(\frac{4x^* + 2y^*}{5}, \frac{2x^* + y^*}{5}\right)\end{aligned} \tag{10.13}$$

Thus giving a reconstruction error of

$$\vec{d} = (|\Delta x|, |\Delta y|) = \left|\left(\frac{4x^* + 2y^*}{5}, \frac{2x^* + y^*}{5}\right) - (x^*, y^*)\right| = \left(\left|\frac{2y^* - x^*}{5}\right|, \left|\frac{2x^* - 4y^*}{5}\right|\right) \tag{10.14}$$

where the absolute value in the second expression is to be understood component-wise.

In order to compute the expected error for each of the dimension, we need only integrate the uniform distribution from $[0, p/2]$ for both dimensions. For the y dimension this becomes

$$\begin{aligned}\mathbb{E}[|\Delta y|] &= \frac{4}{p^2}\int_0^{p/2}\int_0^{p/2} |\Delta y(x', y')|\, dy'\, dx' \\ &= \frac{4}{5p^2}\int_0^{p/2}\int_0^{x'/2} (2x' - 4y')\, dy'\, dx' + \frac{4}{5p^2}\int_0^{p/2}\int_{x'/2}^{p/2} (4y' - x')\, dy'\, dx' \\ &= \frac{4}{5p^2}\int_0^{p/2} \frac{x'^2}{2}\, dx' + \frac{4}{5p^2}\int_0^{p/2}\left(\frac{x'^2}{2} + \frac{p^2}{2} - px'\right) dx' \\ &= \frac{4}{5p^2} \cdot \frac{p^3}{48} + \frac{4}{5p^2} \cdot \frac{7p^3}{48} = \frac{2}{15}p\end{aligned} \tag{10.15}$$

For the square matrix the linking vector always lies on the x-axis since both the seed pixel and its neighbour share the same y coordinate. This simplifies the calculation considerably, since the error in the y dimension will be given by the true position of the hit with respect to the cluster y coordinate, or following convention above, $|\Delta y| = y^*$. The expected error is then

$$\begin{aligned}\mathbb{E}[|\Delta y|] &= \frac{4}{p^2}\int_0^{p/2}\int_0^{p/2} |\Delta y(x', y')|\, dy'\, dx' = \frac{4}{p^2}\int_0^{p/2}\int_0^{p/2} y'\, dy'\, dx' \\ &= \frac{4}{p^2}\int_0^{p/2}\left(\frac{p^2}{8}\right) dx' = \frac{4}{p^2} \cdot \frac{p^3}{16} = \frac{p}{4}\end{aligned} \tag{10.16}$$

The obtained expected error of $\frac{p}{4}$ is almost double that of the staggered matrix arrangement. Though the equivalent estimate for the error is the x dimension is not included here, it is marginally better in square arrangement than in the staggered arrangement, which follows the intuition that in the squared arrangement only the x-residual is minimized. Indeed, these estimates line up well with the observed trend, where the staggered matrix was found to outperform the square arrangement primarily through the the improved reconstruction of the y position of the cluster. This was observed even at very low thresholds of the cluster method, and separately for the window method, explicitly depicted on Figure 10.21.

The staggered matrix arrangement offers an improved spatial resolution by minimizing the loss of information due to charge sharing among neighbours, and optimizing the geometry for hit localization. This effect is especially pronounced along the y-axis, where the pixel arrangement provides more distinct spatial projections, facilitating finer discrimination. The excellent resolution performance, especially under high-threshold operation, highlights the value of geometric optimization as a design lever in future pixel architectures.

10.3 DESY 2023 Testbeam

In addition to SPS, testbeam measurements of different CE-65v2 variants were conducted at the T24 beamline of the DESY II testbeam facility [216] in November 2023. Although this testbeam campaign preceded the SPS testbeam, the observed large residuals limited the usefulness in determining the sensor's intrinsic resolution. This section presents the DESY testbeam setup and results, with particular focus on the influence of biasing voltage on charge sharing.

10.3.1 Setup

Testbeam Facility

The 293 m DESY II synchrotron serves as an injector for PETRA III [224], providing electrons and positrons with beam energies of up to 7 GeV. Beam generation for the testbeam facilities proceeds parasitically by inserting a carbon fibre target into the primary DESY II beam, causing the emission of energetic photons via bremsstrahlung. The photons are subsequently converted to electron-positron pairs by striking a secondary target consisting of Cu or Al. The resulting beam passes through a set of dipole magnets located 60 cm downstream of the secondary target, allowing for the selection of particle parity and momentum through adjustment of the magnetic field. The selected particles, either electrons or positrons, are then collimated and served to the testbeam areas at user-selected energies ranging from 1-6 GeV. For all results presented in this thesis, a 4 GeV/c electron beam was used.

Beam Telescope

The same ALPIDE-based telescope was used for the DESY II testbeam. A similar arrangement was adopted, with three reference planes positioned upstream and downstream of the DUT and the trigger, respectively. However, the distance between the upstream and downstream telescope arms was increased to 100 mm, increasing the separation by 25 mm with respect to the SPS setup. The telescope used at the DESY testbeam is depicted on Figure 10.27. The telescope resolution was estimated to be 3.2 µm using the telescope optimizer tool [221], which is likely a significant underestimate based on the observed residuals, as detailed in the following Section.

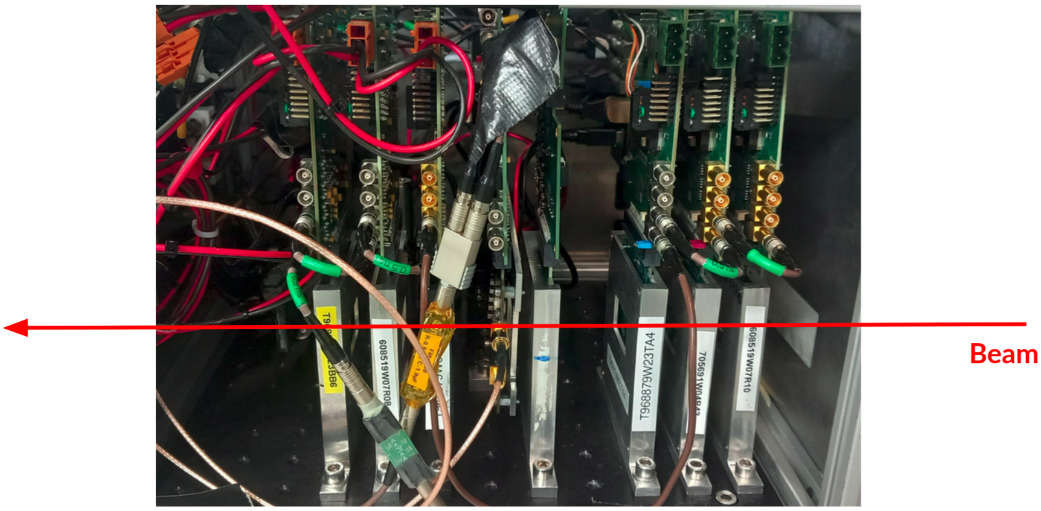


Figure 10.27: Photograph of the CE-65v2 telescope at the DESY testbeam. The DUT can be identified by the blue marking on the unlabelled metallic holder. The beam direction is indicated by the red arrow. During data taking the setup was covered to prevent light exposure of the detector planes.

10.3.2 Results

The measurements conducted at the DESY testbeam were analogous to those conducted at the SPS testbeam. The same analysis strategy was adopted, and thus the reader is referred to Section 10.2.2. The primary difference was the adoption of the GBL track reconstruction algorithm, in order to account for multiple scattering.

The normalized distributions of the cluster and seed charges for the window clusterisation method at a seed threshold of 100 e^- are shown in Figures 10.28 and 10.29, respectively. The distributions behave as expected, with the collected cluster signal being virtually identical for all four of the chips. The seed charge distribution reveals a higher MPV for the two Modified with Gap chips, corresponding to the lower charge sharing with respect to the Standard process. The measured distributions match those of the SPS testbeam, depicted on Figure 10.7, very well, which is expected as the deposited charge by MIPs is defined by the sensor properties, and in particular the thickness of the epitaxial layer [212]. Indeed, the apparent saturation of the sensors is once again evident at ~ 1700 e^- in the seed distribution for all four chips.

Large Residuals

The measured spatial resolutions, however, show little agreement with those measured at the SPS testbeam. Figure 10.30 presents the x-residuals for the 15 µm and 22.5 µm Standard and Modified with Gap CE-65v2 chips. The distributions for all four chips show large Gaussians about a central value, even for the Modified with Gap chips, where the convolution with a box function is expected. When comparing the measured residuals with those of Figure 10.12, a clear trend emerges that the standard deviation is around two times larger for each of the chips. Indeed, even when accounting for the worse telescope resolution expected with the lower energy electrons, the spatial resolution is still substantially higher for all four of the chips.

Telescope Resolution Estimate

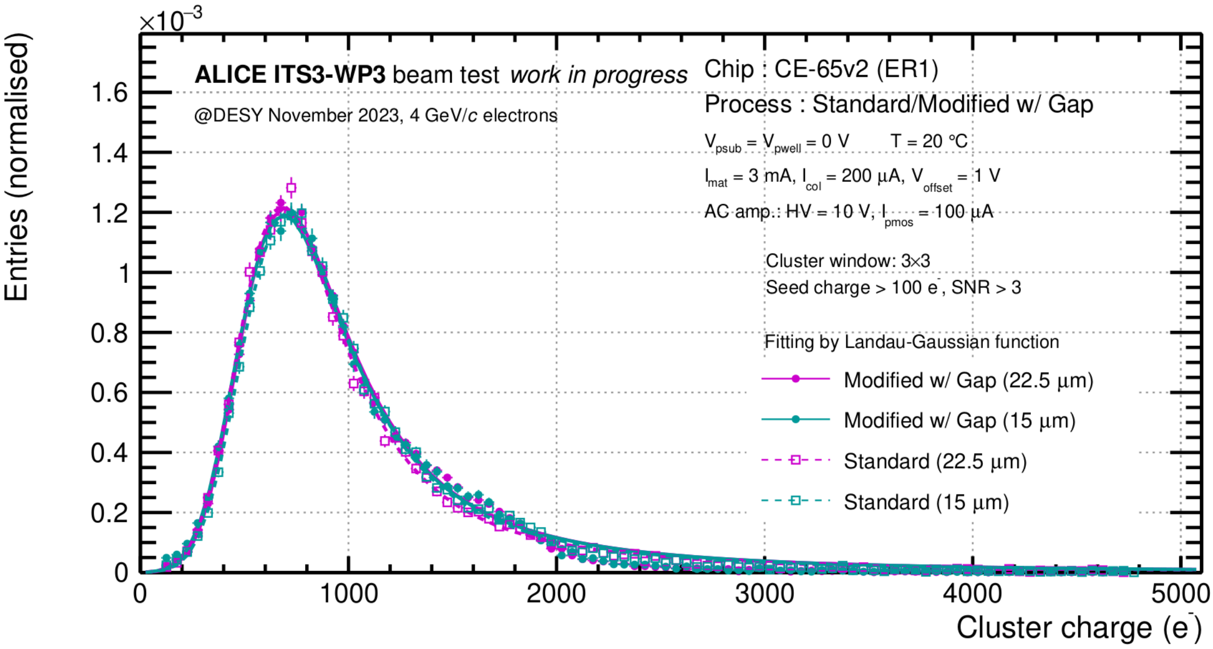


Figure 10.28: Normalised distribution of total cluster charge at a seed threshold of 100 e$^-$ for the 4 considered CE-65v2 chips at the DESY testbeam. All four distributions were fit by a Landau-Gaussian convolution, as defined in Equation 10.9. The window method was used for reconstruction.

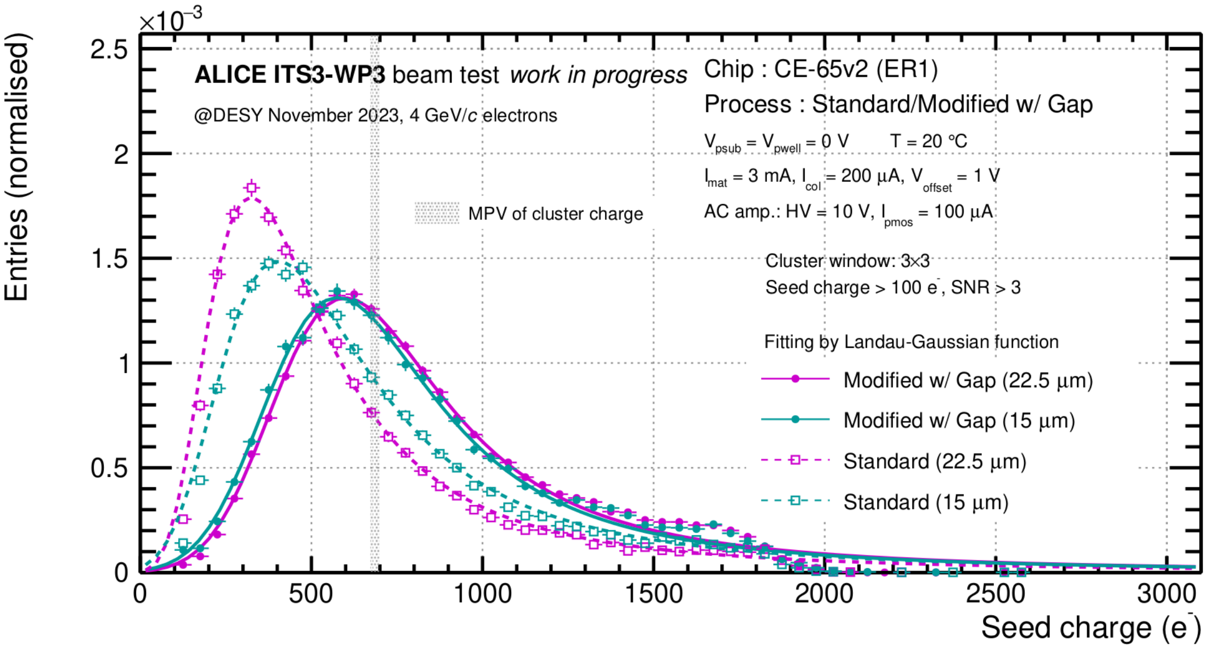


Figure 10.29: Normalised distribution of seed charge at a seed threshold of 100 e$^-$ for the 4 considered CE-65v2 chips at the DESY testbeam. All four distributions were fit by a Landau-Gaussian convolution, as defined in Equation 10.9. The MPV of the cluster charge from Figure 10.6 is overlaid for reference. The window method was used for reconstruction.

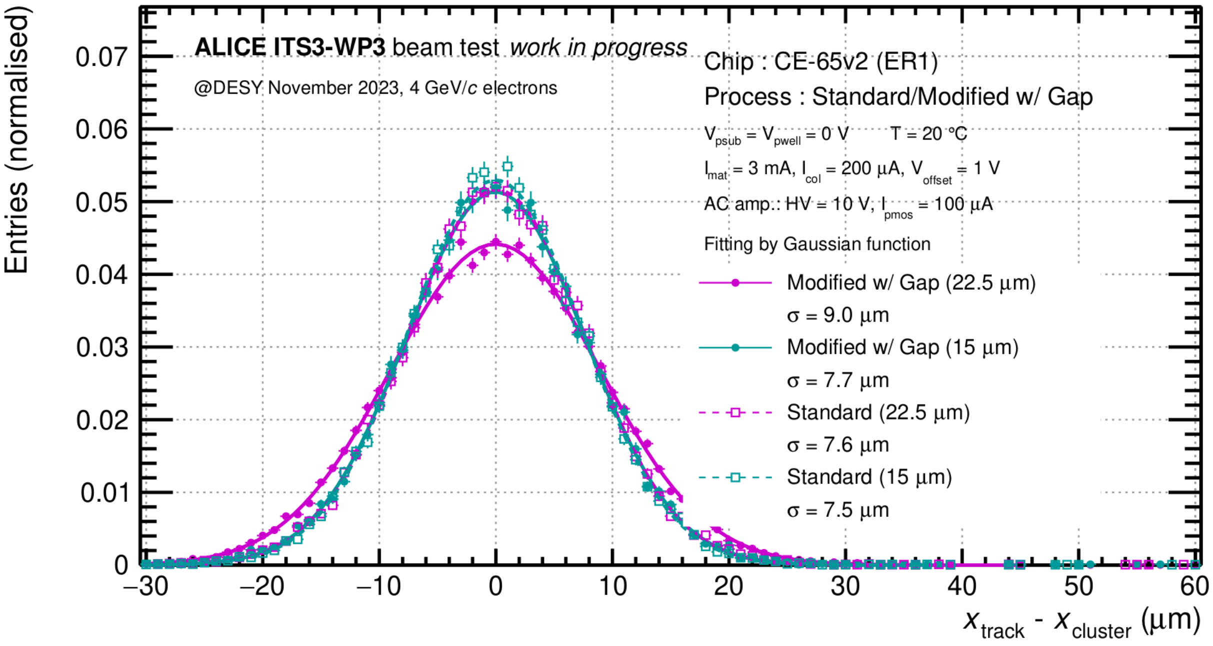


Figure 10.30: Distribution of x-residuals at a seed threshold of 100 e$^-$ for the 4 considered CE-65v2 chips at the DESY testbeam. Gaussians were fit to each of the chips to extract the corresponding σ. The window method was used for reconstruction.

An attempt at estimating the telescope resolution of the beam telescope at the DESY testbeam was made by noting that

$$\sigma_{spatial} = \sqrt{(\frac{\sigma_{\Delta x}^{DESY} + \sigma_{\Delta y}^{DESY}}{2})^2 + (\sigma_{tel.}^{DESY})^2} \tag{10.17}$$

can be solved for the telescope resolution at DESY, while using the measured spatial resolution at the SPS testbeam

$$\sigma_{tel.}^{DESY} = \sqrt{\left(\frac{\sigma_{\Delta x}^{DESY} + \sigma_{\Delta y}^{DESY}}{2}\right)^2 - \left(\sigma_{spatial}^{SPS}\right)^2} \tag{10.18}$$

In this way, the DESY telesope resolution was estimated at $\sigma_{tel.}^{DESY} = 7.27 \pm 0.16$ µm. While such a large telescope resolution limits the usefulness of the DESY testbeam measurements for the computation of the spatial resolution of the CE-65v2 variants, the large statistics obtained at the DESY testbeam allowed the study of the evolution of the charge sharing properties of the CE-65v2 variants with respect to biasing voltage. These measurements, coupled with those presented in Section 10.2.5, enabled the study of charge sharing with respect to all design choices of the CE-65v2 chip.

Multiple Scattering

The estimated telescope resolution is over three times that observed at the SPS testbeam. While some degradation in the telescope resolution is expected due to the lower energy electron beam resulting in considerably more multiple scattering, the observed effect is expected to be smaller. A key difference between the SPS and DESY testbeam campaigns is that the distance between the DUT and the last reference plane of the upstream arm is doubled, as described in Section 10.3.1. The increased distance between DUT and the upstream arm is known to significantly degrade the telescope resolution [225]. Although the telescope optimiser in principle accounts for such geometry, additional material (e.g. the cloth covering the telescope) may have enhanced scattering effects more strongly during the DESY testbeam than at SPS, where the high energy hadron beam is less sensitive to such perturbations.

Charge Sharing

The accumulated charge ratios, as defined in Section 10.1.1, were measured for the four CE-65v2 chips at biasing voltages of $[10, 4, 2, 0]$ V. Unsurprisingly, the accumulated charge ratio for the four chips almost exactly mirrored the accumulated charge ratio measured at the SPS testbeam at a biasing voltage of 10 V, depicted on Figures 10.18 and 10.19 for the Modified with Gap and Standard process chips, respectively. The average charge fraction collected by the central pixel is virtually the same as for the SPS testbeam for all four of the chips. For the Standard process chips this corresponds to 66.0% and 58.6% for the 15 µm and 22.5 µm pitch chips, respectively. For the Modified with Gap process chips this corresponds to 84.2% and 86.7% for the 15 µm and 22.5 µm pitch chips, respectively.

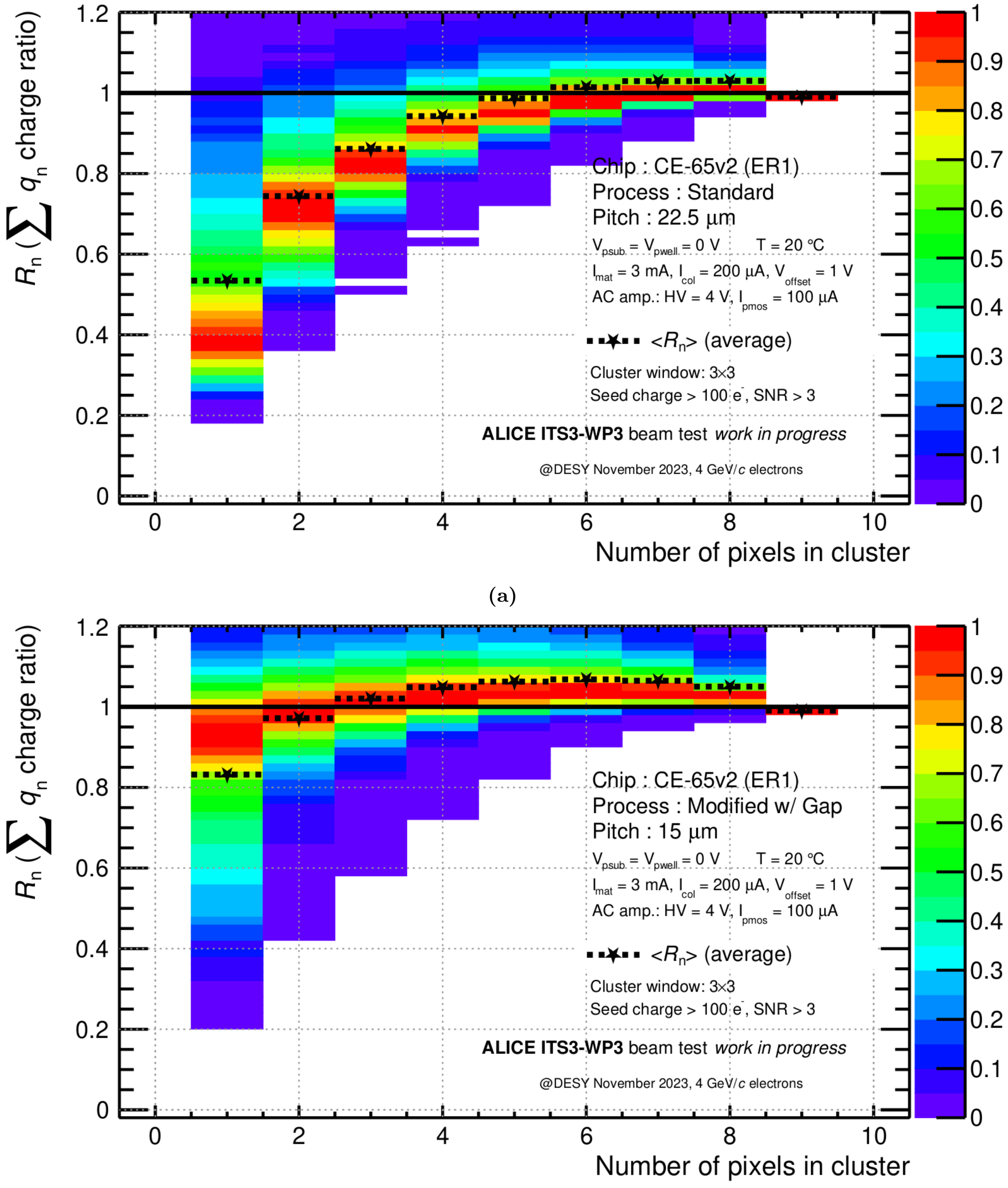
Figure 10.31a depicts the measured accumulated charge ratio for the 22.5 µm Standard process chip at a biasing voltage of 4V. A sizeable decrease in the accumulated charge in the first couple bins can be observed, which diminishes in magnitude as the cluster size increases. In particular, the average charge fraction collected by the central pixel is now 53.4%, marking a 9% decrease in the collected charge. This can be associated with the development of the depletion region as the biasing voltage is increased, evident also in Figure 9.11 from the lab tests.

Figure 10.31b depicts the equivalent result at 4 V for the 15 µm Modified with Gap chip. For the Modified with Gap process chip a considerably milder decrease in the average charge fraction can be observed to 83.2%. For cluster sizes of 2 and above the decrease in the accumulated charge ratio is negligible, showing that, despite the larger capacitance change observed when moving from 4 V to 10 V in Figure 9.11, the charge sharing behaviour of the Modified with Gap process shows little sensitivity to the biasing voltage at this range. This lines up with the expectation that this capacitance change does not result from the vertical depletion of the epitaxial layer, but rather with the depletion of the low dose n-type implant around the collection electrode.

Figure 10.32a depicts the measured accumulated charge ratio for the 22.5 µm Standard process chip at a lower biasing voltage of 2 V. A continuation of the trend observed for the biasing voltage of 4 V can be observed, with the charge ratio again decreasing considerably for all low cluster size bins. Indeed, the decrease in the charge ratio is noticeable until a cluster size of 5. The average charge fraction collected in the central pixel decreases to 49.7%, which is smaller than the decrease observed for a biasing voltage of 10 V, although still comparable.

The equivalent result for the 15 µm Modified with Gap chip, depicted on Figure 10.32b, again shows a considerably milder decrease than its Standard process counterparts. Indeed, above a cluster size of 2, the change in the accumulated charge fraction is hardly perceptible. The average charge fraction collected by the central pixel decreases to 81.7%, which although perceptible, does not compare to the biasing voltage dependence observed for the Standard process.

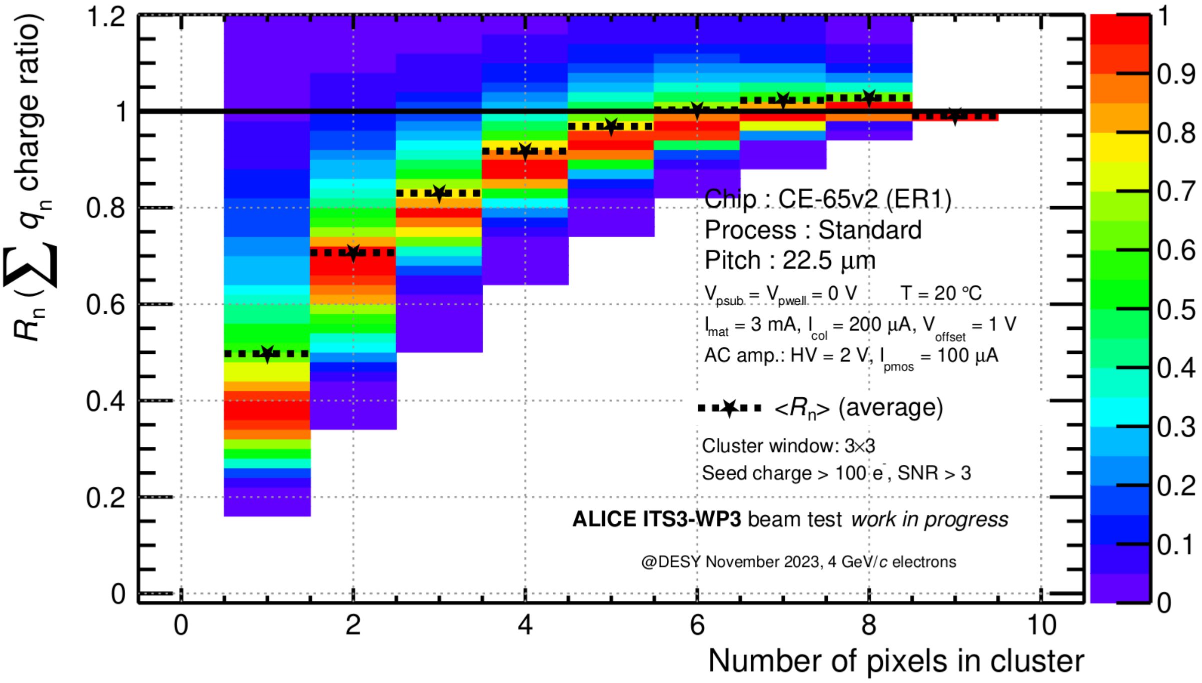
Despite the challenges posed by multiple scattering and limited telescope resolution, the DESY testbeam confirmed key trends also observed at the SPS and provided further insight into the charge sharing behaviour of the CE-65v2 chips. The results illustrate that the biasing voltage can serve as an effective handle to tune charge sharing in the Standard process chips,



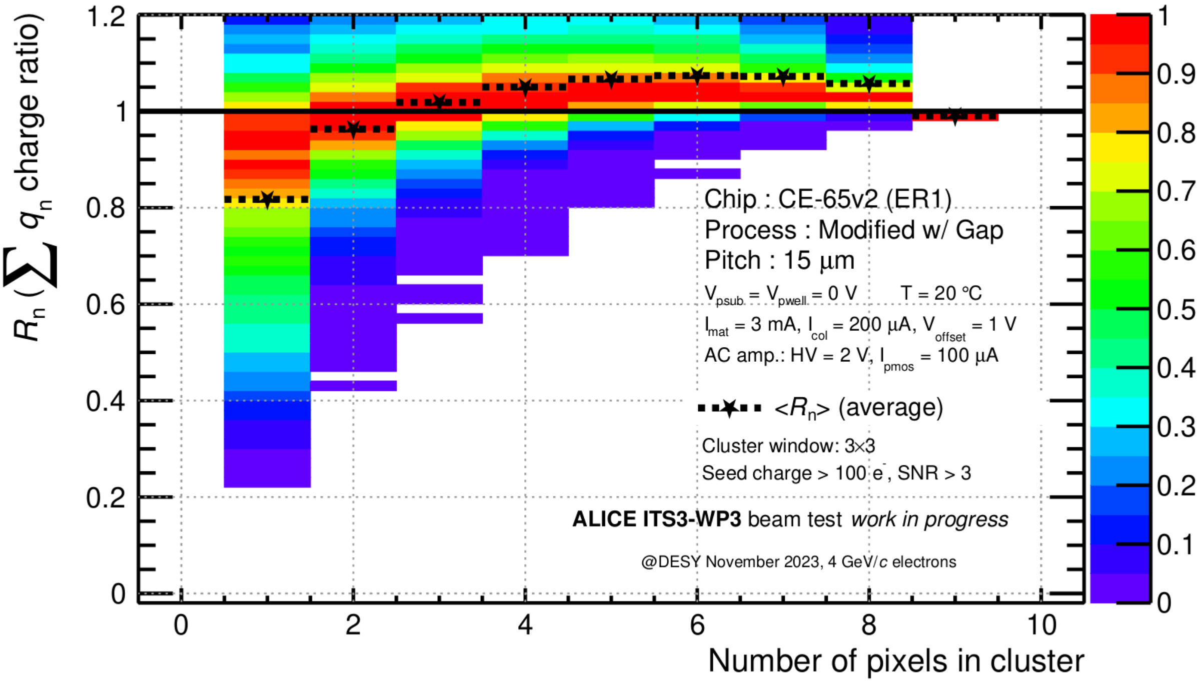
(a)

(b)

Figure 10.31: Accumulated charge ratio as a function of number of pixels in a cluster at a biasing voltage of 4 V for the CE-65v2 chip with a 22.5 µm pitch in the Standard process (a) and a 15 µm pitch in the Modified with Gap process (b).



(a)



(b)

Figure 10.32: Accumulated charge ratio as a function of number of pixels in a cluster at a biasing voltage of 2 V for the CE-65v2 chip with a 22.5 µm pitch in the Standard process (a) and a 15 µm pitch in the Modified with Gap process (b).

where the seed pixel collects between $\sim$ 43% at 0 V and $\sim$ 60% at full depletion. On the other hand, the Modified with Gap process exhibits a much weaker dependence on bias voltage, limiting its tunability. In particular, it was observed that even at 2 V, an average of 95+% of the signal is carried by the first two pixels in a cluster. Although increased charge sharing can be achieved at even lower voltages, this comes at the cost of rapidly rising sensor capacitance as the n-type implant around the collection electrode is still undepleted in this biasing regime. Nevertheless, the measurements confirm that the Modified with Gap chips can be operated at high biasing voltages with minimal loss of charge sharing, offering the advantage of low capacitance and improved noise performance.

10.4 Conclusions

A comprehensive comparison of the Standard and Modified with Gap process variations, pixel pitch (15 µm, 22.5 µm), and matrix geometry (square vs staggered), was performed during two testbeam campaigns using the CE-65v2 chip. Particular emphasis was placed on the charge sharing properties, and their ramifications on the hit-detection efficiency and spatial resolution, across the different design variants.

An excellent resolution was obtained in large-matrix 65 nm CMOS test structures during a test beam campaign at the CERN SPS. In-pixel studies demonstrated a uniform spatial resolution across the Standard process CE-65v2 pixel, highlighting the accurate position reconstruction for events far from the pixel centre. The sub 3 µm spatial resolution obtained in the Standard process for both the 15 µm and 22.5 µm pitch satisfies FCC-ee requirements [226], and allows tradeoffs in-pixel pitch with respect to power consumption, readout-rate, and manufacturing-ease. The 15 µm Modified with Gap process, while not quite reaching sub 3 µm spatial resolution, is likely to meet the requirement with a modest reduction in pixel pitch, retaining all benefits associated with the process modification. This enables a wide range of applications for 65 nm TPSCo MAPS for both current, and next-generation detector systems.

However, the spatial resolution was observed to quickly degrade when Standard process chips are operated outside the low electron threshold regime, making the process particularly sensitive to noise and radiation defects, including charge trapping. The Modified with Gap process, on the other hand, displays a wide operating range, with an efficiency of over 99% being achievable up to $\sim$ 180 e^- for both the 15 µm and 22.5 µm chips. While the wider operating margin and faster charge collection make the process more suited for high radiation environments, in pixel studies demonstrated a considerable degradation in the spatial resolution when moving away from the pixel centre, with the observed spatial resolution being as high as 6 µm at the pixel corners.

Adopting a staggered matrix arrangement resulted in an improved reconstruction of the y position, particularly for the Standard process chips. While gains in efficiency were marginal, the spatial resolution in the range 150-250 e^- improved considerably for the 22.5 µm Standard process chip, with a 16% improvement with respect to the square arrangement at 210 e^-.

The consistently superior spatial resolution of the Standard process can be directly attributed to enhanced charge sharing, which was further investigated across process modifications and pixel pitches. The 22.5 µm Standard process chips exhibited an average charge fraction collected by the seed pixel of 58.6%, which is substantially lower than its Modified with Gap counterpart at 86.8%. An inverse relationship between pixel pitch and charge sharing was observed between both process modifications, with a larger pitch leading to more charge sharing for the Standard process. While the biasing voltage was successfully used as a lever to tune charge sharing in the Standard process chips, it was of limited effect for their Modified with Gap counterparts.

Digitisation studies demonstrated that the spatial resolution in Modified with Gap process only marginally benefits from more sophisticated reconstruction methods, with the resolution obtained using the binary method 3.22 ± 0.02 µm improving only to 3.09 ± 0.02 µm when the window method is used. In contrast, the Standard process variants, with their high charge sharing, benefitted considerably more, with the spatial resolution improving from 3.74 ± 0.02 µm to 2.02 ± 0.02 µm under the same change in clusterisation approach. However, the full improvement in the reconstruction of the cluster was already observed at a 6-bit encoding, suggesting that a significant reduction in data rate is possible, without compromising reconstruction quality.

In summary, the characterisation of the CE-65v2 chip has supplemented the APTS [197] and DPTS [198] studies in the validation of the 65 nm TPSCo process as a candidate technology for advanced particle detection applications, including the ALICE ITS3 upgrade. The demonstrated spatial resolution, and its dependence on process modifications, pixel pitch, and matrix geometry, provides valuable input for guiding future design choices in next-generation tracking detectors for high-energy physics. Ongoing investigations into radiation tolerance and the Modified process variants will extend these findings and help assess the applicability of the adopted technologies under diverse radiation conditions.

Part IV

Conclusion and Outlook

11 Conclusion

Future lepton colliders such as the FCC-ee promise a rich physics programme in a clean environment at unparalleled statistics. From electroweak measurements at the Z-pole to precision studies of the Higgs boson, the physics programme at the FCC-ee will probe the Standard Model with unprecedented scrutiny, offering a unique window into potential new physics. Maximally exploiting the physics potential of future colliders hinges on the precise reconstruction of final states, and in particular of the partons from which it originates. This thesis reports on two complementary pillars of final state reconstruction at future colliders: jet flavour tagging and charged particle tracking. Together they exemplify how advances in both algorithms and hardware domains will be instrumental for achieving the ambitious goals set out for the FCC-ee.

Jet Flavour Tagging at the FCC-ee

Part II revolves around exploring the prospects of flavour tagging at the FCC-ee, with particular focus on the identification of strange quark initiated jets, a task that has long remained elusive due the absence of long-lived hadrons. To this end a multiclassifier neural network using a transformer-based architecture, dubbed *DeepJetTransformer*, was used to discriminate jets using Delphes fast simulation to simulate the IDEA detector concept. By combining the representational power of the self-attention mechanism, able to capture subtle correlations between jet constituents, with novel vertexing strategies at the FCC-ee, the tagger achieved state-of-the-art performance in a wide variety of classification tasks. Across both $Z \to q\bar{q}$ and $ZH(\to q\bar{q})$ decays, the model consistently achieved high performance, with efficiencies above $\epsilon_{sig} \sim 87\%$ at a $\epsilon_{bkg} \sim 10\%$ background efficiency for b vs c and c vs s discrimination. Kaon identification proved to be instrumental for s-jet discrimination, where u- and d-jets were found to dominate the background. Charged kaon identification was emulated through the inclusion of a $K^{\pm}$ ID flag sampled from a uniform distribution at varying efficiencies. By contrast, neutral kaon identification was treated more realistically, with K^0_S candidates included explicitly through the reconstruction of V^0 vertices during vertexing. Following the evaluation of tagging performance, and its dependence on varying levels of particle identification, subsequent permutation feature importance studies demonstrated the dominant impact of track and secondary vertex variables to heavy flavour tagging, while further highlighting the importance of particle identification for s-jet discrimination.

DeepJetTransformer is part of a wider shift towards transformer architectures in the context of jet flavour tagging. Their rapid adoption is motivated by their improved representational power and computational efficiency. The detector concept landscape at the FCC-ee is rapidly evolving, with optimisations ranging from adjustments in the subdetector geometry to improvements in the performance of individual detector modules, such as the silicon pixel detectors presented in Part III of this thesis. With a computational complexity of just 19.7 MFLOPs, DeepJetTransformer offers a compelling avenue for the iterative studies necessary to quantify the ramifications of detector design choices. The observed necessity of kaon identification in isolating strange jets reinforces a growing body of evidence highlighting its critical role at the FCC-ee. These findings align with the increasing emphasis on the inclusion of particle identification in the detector concepts presented in Chapter 3, including for instance the potential adoption of a RICH detector in the CLD design.

Future work on the DeepJetTransformer algorithm may be centered around improvements to the input feature set. While there is significant room for optimisation of the architecture and hyperparameter tuning, including adjustments to layer parameters and the network structure, the most impactful gains would stem from enhancing the realism and relevance of the input feature set. The current feature set is likely far from optimal, partially due to degraded vertexing performance under the assumed innermost vertex layer radius of 1.7 mm, relative to the present baseline of 1.37 mm. A natural starting point would be to incorporate an updated description of the IDEA detector concept, as well as a substantially larger input dataset. With $\sim 10^6$ jets, comparable to the number of trainable parameters, the current input dataset effectively constitutes a rough lower bound for model training. Further gains could be achieved by extending the input feature set to include additional features, including the full covariance matrix. Moreover, the permutation feature importance studies highlighted overlap in the current input feature set, which could be exploited to reduce computational complexity. The baseline $K^{\pm}/\pi^{\pm}$ discrimination is overly conservative with respect to the particle identification performance forseen at the FCC-ee. The inclusion of more realistic PID variables, such as the time-of-flight mass $m_{t.o.f.}$ and the number of ionisation clusters dN/dx, would not only add realism, but significantly improve the s-jet tagging performance achievable with DeepJetTransformer. Finally, the subdivision of jet flavours into categories with distinct signatures, such as hadronic and semi-leptonic B hadron decays, or $g \to b\bar{b}$ splittings that may not resemble the typical radiation pattern of a gluon jet, is likely to offer further improvements in performance. Nevertheless, the optimisation of present algorithms is only of limited interest given the FCC-ee's distant timeline. The primary motivation lies in assessing the physics potential based on specific design choices for detector systems and the FCC-ee accelerator as a whole.

Monolithic Active Pixel Sensors for Future Colliders

Monolithic Active Pixel Sensors have emerged as the enabling technology for vertexing systems at the FCC-ee. By combining the sensing layer and the readout circuitry onto the same silicon die, MAPS offer a reduced material budget, increased spatial resolution, and decreased power consumption, albeit at the cost of a reduced radiation tolerance. MAPS are thus uniquely suited to environments where high granularity is the primary constraint, such as in heavy-ion collisions, or future lepton colliders such as the FCC-ee. Part III reports on the characterisation of a test structure produced in the 65nm TPSCo CMOS imaging process: the CE-65. The CE-65v1 and CE-65v2 chips were produced in a wide variety of variants targetting four exploration axes: amplification scheme, process modification, pitch variation, and matrix geometry. During the initial lab characterisation of the CE-65 family of chips at the University of Zurich, the chips exhibited low noise performance in the range of 15 - 25 e^- across a variety of different pixel pitches and process modifications. Subsequent gain calibration and uniformity studies using an ^{55}Fe source demonstrated a high-gain uniformity of $\mathcal{O}(4\%)$ for all considered chips, and a relative gain of 1:4:3 for the 25 µm Standard process CE-65v1 chip and 1:3:4 for its 15 µm Modified process counterpart, in agreement with simulation. The AC-coupled preamplifier displayed the highest charge collection efficiency ($\sim$ 97%) and best energy resolution ($\sim$ 7%), vindicating its adoption as the sole preamplifier architecture in the CE-65v2 chip.

The subsequent characterisation of the CE-65v2 at testbeams at CERN SPS and DESY provided a comprehensive comparison of the Standard and Modified with Gap process variations, pixel pitch (15 µm, 22.5 µm), and matrix geometry (square vs staggered). Substantially higher charge sharing was observed for the Standard process chips, allowing a precise reconstruction of impinging particles. A sub 3 µm spatial resolution was achieved for the both 15 µm and 22.5 µm pitch CE-65v2 chips, satisfying FCC-ee requirements. Reducing the precision of the collected charge to a 6-bit encoding preserved the reconstruction precision, offering potential gains in data volume. By contrast, the Modified with Gap process chips display much less charge sharing, and thus only marginally benefit from sophisticated position reconstruction methods. Despite displaying a consistently worse spatial resolution than the Standard process chips, the 15 µm Modified with Gap process chip still achieves a $\sim$ 3 µm spatial resolution, and a considerably wider operational margin, maintaining an efficiency of over 99% up to $\sim$ 180 e^-. The spatial resolution was further improved through the adoption of a staggered matrix geometry for all considered chip variants, with improvements particularly pronounced for the Standard process chips at moderate thresholds of 150-250 e^-. The improvement in the spatial resolution is primarily attributed to the reduced charge sharing among neighbours, and the beneficial geometry of the staggered arrangement. The adoption of a staggered matrix geometry only marginally impacted the hit-detection efficiency of the CE-65v2 chip.

The CE-65 characterisation presented in this thesis complements existing APTS [197] and DPTS [198] studies, and further consolidates the 65nm TPSCo imaging process as a candidate technology for high-resolution silicon tracking. The systematic study of the aforementioned *exploration axes* provides valuable input for future MAPS detectors development. The recently formed Optimized CMOS Technology fOr Precision in Ultra-thin Silicon (OCTOPUS) project [227] aims to develop a full-size sensor demonstrator satisfying the stringent requirements of future lepton collider vertex detectors. In line with the presented results, two promising avenues have emerged to satisfy the multifaceted requirements of future vertex detectors, and in particular the sub 3 µm spatial resolution. The first is a small-pitch sensor produced in the Modified with Gap process with a binary readout. While the 15 µm CE-65v2 chip does not quite satisfy the spatial resolution requirement, a modest reduction in pixel pitch is likely to achieve a sub 3 µm spatial resolution with a binary readout. Such an approach would entail a fast charge collection and a high hit-detection efficiency at moderate electron thresholds, thereby enhancing radiation hardness. The small pixel pitch would, however, introduce challenges regarding the miniaturisation of the in-pixel electronics, and would limit the achievable spatial resolution due to the binary readout scheme. The second avenue would be a considerably larger pitch Standard process chip with a charge-sensitive readout. As demonstrated in this thesis, a 22.5 µm Standard process chip achieves a $\sim$ 2 µm resolution when operated at low thresholds through the exploitation of significant charge sharing during the position reconstruction. Novel readout strategies are prerequisite to unlocking combinations of design parameters explored with the CE-65(v2) chips, enabling a maximal single point resolution whilst not compromising readout speed or power efficiency. The newly developed SPARC chip [228], for instance, implements an asynchronous matrix readout architecture based on Fixed-Priority Arbiters (FPAs), allowing large groups of pixels to be readout in a fast and power-efficient way. While the larger pitch and improved spatial resolution with respect to the Modified with Gap approach are compelling, it remains unclear if the Standard process can simultaneously satisfy the remaining requirements, and in particular the radiation tolerance, due to its slow, diffusion-dominated charge collection, and rapid decrease in hit-detection efficiency when operated at higher thresholds. Investigations into the radiation tolerance of the CE-65v2 chip variants are ongoing at the time of writing, aiming to extend the presented findings, and to clarify whether the Standard process remains a viable candidate under the conditions expected at future lepton colliders.

References

[1] A. De Moor, "Deep learning for jet algorithms". PhD thesis, Vrije Universiteit Brussel, Belgium, 2024.

[2] F. Blekman et al., "Tagging more quark jet flavours at FCC-ee at 91 GeV with a transformer-based neural network", *Eur. Phys. J. C* **85** (2025), no. 2, 165, `doi:10.1140/epjc/s10052-025-13785-y`, `arXiv:2406.08590`.

[3] D. Dannheim et al., "Corryvreckan: A Modular 4D Track Reconstruction and Analysis Software for Test Beam Data", *JINST* **16** (2021), no. 03, P03008, `doi:10.1088/1748-0221/16/03/P03008`, `arXiv:2011.12730`.

[4] **ALICE** Collaboration, "Characterisation of analogue MAPS produced in the 65 nm TPSCo process", *JINST* **20** (2025), no. 01, C01019, `doi:10.1088/1748-0221/20/01/C01019`, `arXiv:2411.08740`.

[5] **ATLAS** Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, `doi:10.1016/j.physletb.2012.08.020`, `arXiv:1207.7214`.

[6] **FCC** Collaboration, "FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2", *Eur. Phys. J. ST* **228** (2019), no. 2, 261–623, `doi:10.1140/epjst/e2019-900045-4`.

[7] "International Linear Collider reference design report. 1: Executive summary. 2: Physics at the ILC. 3: Accelerator. 4: Detectors", `doi:10.2172/929487`.

[8] **CEPC Study** Group, "CEPC Conceptual Design Report: Volume 1 - Accelerator", `arXiv:1809.00285`.

[9] M. D. Schwartz, "Quantum Field Theory and the Standard Model". Cambridge University Press, 3, 2014. ISBN 978-1-107-03473-0, 978-1-107-03473-0.

[10] D. Galbraith and C. Burgard, "Diagram of the Standard Model of particle physics, created at the CERN Webfest 2012". `https://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model`, 2012. [Online; accessed 11-Mar-2021].

[11] J. Rohlf, "Modern Physics from alpha to Z0". Wiley, 1994. ISBN 9780471572701.

[12] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346, `doi:10.1103/PhysRevLett.30.1343`.

[13] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, `doi:10.1103/PhysRevLett.19.1264`.

[14] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, `doi:10.1016/0029-5582(61)90469-2`.

[15] A. Salam, “Weak and Electromagnetic Interactions”, *Conf. Proc. C* **680519** (1968) 367–377, `doi:10.1142/9789812795915_0034`.

[16] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10** (1963) 531–533, `doi:10.1103/PhysRevLett.10.531`.

[17] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, *Prog. Theor. Phys.* **49** (1973) 652–657, `doi:10.1143/PTP.49.652`.

[18] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508–509, `doi:10.1103/PhysRevLett.13.508`.

[19] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321–323, `doi:10.1103/PhysRevLett.13.321`.

[20] M. B. Green, J. H. Schwarz, and E. Witten, “Superstring Theory. Vol. 1: Introduction”. Cambridge Monographs on Mathematical Physics. 7, 1988. ISBN 978-0-521-35752-4.

[21] M. B. Green, J. H. Schwarz, and E. Witten, “Superstring Theory. Vol. 2: Loop Amplitudes, Anomalies and Phenomenology”. 7, 1988. ISBN 978-0-521-35753-1.

[22] V. C. Rubin, W. K. Ford, Jr., and N. Thonnard, “Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa through Sc”, *Astrophys. J. Lett.* **225** (1978) L107–L111, `doi:10.1086/182804`.

[23] K. G. Begeman, A. H. Broeils, and R. H. Sanders, “Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics”, *Mon. Not. Roy. Astron. Soc.* **249** (1991) 523, `doi:10.1093/mnras/249.3.523`.

[24] M. C. Gonzalez-Garcia and M. Maltoni, “Phenomenology with Massive Neutrinos”, *Phys. Rept.* **460** (2008) 1–129, `doi:10.1016/j.physrep.2007.12.004`, `arXiv:0704.1800`.

[25] P. Fisher, B. Kayser, and K. S. McFarland, “Neutrino Mass and Oscillation”, *Ann. Rev. Nucl. Part. Sci.* **49** (1999) 481–528, `doi:10.1146/annurev.nucl.49.1.481`, `arXiv:hep-ph/9906244`.

[26] P. F. de Salas et al., “2020 global reassessment of the neutrino oscillation picture”, *JHEP* **02** (2021) 071, `doi:10.1007/JHEP02(2021)071`, `arXiv:2006.11237`.

[27] “A Multi-TeV Linear Collider Based on CLIC Technology: CLIC Conceptual Design Report”, `doi:10.5170/CERN-2012-007`.

[28] **FCC** Collaboration, “Future Circular Collider Feasibility Study Report: Volume 2, Accelerators, Technical Infrastructure and Safety”, `doi:10.17181/CERN.EBAY.7W4X`, `arXiv:2505.00274`.

[29] **FCC** Collaboration, “Future Circular Collider Feasibility Study Report: Volume 1, Physics, Experiments, Detectors”, `doi:10.17181/CERN.9DKX.TDH9`, `arXiv:2505.00272`.

[30] C. Accettura et al., “Towards a muon collider”, *Eur. Phys. J. C* **83** (2023), no. 9, 864, `doi:10.1140/epjc/s10052-023-11889-x`, `arXiv:2303.08533`. [Erratum: Eur.Phys.J.C 84, 36 (2024)].

[31] W. Herr and B. Muratori, “Concept of luminosity”, `doi:10.5170/CERN-2006-002.361`.

[32] D. d’Enterria, A. Poldaru, and G. Wojcik, “Measuring the electron Yukawa coupling via

resonant s-channel Higgs production at FCC-ee", *Eur. Phys. J. Plus* **137** (2022), no. 2, 201, `doi:10.1140/epjp/s13360-021-02204-2`, `arXiv:2107.02686`.

[33] **ALEPH** Collaboration, "A Direct measurement of |V(cs)| in hadronic W decays using a charm tag", *Phys. Lett. B* **465** (1999) 349–362, `doi:10.1016/S0370-2693(99)01088-6`.

[34] **OPAL** Collaboration, "A Measurement of the rate of charm production in W decays", *Phys. Lett. B* **490** (2000) 71–86, `doi:10.1016/S0370-2693(00)00971-0`, `arXiv:hep-ex/0009020`.

[35] **DELPHI** Collaboration, "Measurement of |V(cs)| using W decays at LEP-2", *Phys. Lett. B* **439** (1998) 209–224, `doi:10.1016/S0370-2693(98)01061-2`.

[36] A. Freitas et al., "Theoretical uncertainties for electroweak and Higgs-boson precision measurements at FCC-ee", `arXiv:1906.05379`.

[37] P. Skands and D. d'Enterria, "QCD and $\gamma\gamma$ studies at FCC-ee", *PoS* **ICHEP2016** (2017) 1156, `doi:10.22323/1.282.1156`, `arXiv:1610.06254`.

[38] **FCC** Collaboration, "QCD at the Future Circular e^+e^- Collider", in *29th International Workshop on Deep-Inelastic Scattering and Related Subjects.* 8, 2022. `arXiv:2208.08250`.

[39] G. Dissertori et al., "Determination of the strong coupling constant using matched NNLO+NLLA predictions for hadronic event shapes in e+e- annihilations", *JHEP* **08** (2009) 036, `doi:10.1088/1126-6708/2009/08/036`, `arXiv:0906.3436`.

[40] A. Blondel and P. Janot, "FCC-ee overview: new opportunities create new challenges", *Eur. Phys. J. Plus* **137** (2022), no. 1, 92, `doi:10.1140/epjp/s13360-021-02154-9`, `arXiv:2106.13885`.

[41] P. Azzi and E. Perez, "Exploring requirements and detector solutions for FCC-ee", *Eur. Phys. J. Plus* **136** (2021), no. 11, 1195, `doi:10.1140/epjp/s13360-021-02141-0`, `arXiv:2107.04509`.

[42] **RD-FA** Collaboration, "IDEA: A detector concept for future leptonic colliders", *Nuovo Cim. C* **43** (2020), no. 2-3, 27, `doi:10.1393/ncc/i2020-20027-2`.

[43] G. Gaudio, "The IDEA detector concept for FCCee", *PoS* **ICHEP2022** (11, 2022) 337, `doi:10.22323/1.414.0337`.

[44] **IDEA Study** Group, "The IDEA detector concept for FCC-ee", `arXiv:2502.21223`.

[45] L. Pancheri et al., "A 110 nm CMOS process for fully-depleted pixel sensors", *JINST* **14** (2019), no. 06, C06016, `doi:10.1088/1748-0221/14/06/C06016`.

[46] ALICE Collaboration, "Technical Design Report for the ALICE Inner Tracking System 3 (ITS3): A Bent Wafer-Scale Monolithic Pixel Detector", Technical Design Report CERN-LHCC-2024-003, ALICE-TDR-021, CERN, Geneva, 2024.

[47] **FCC** Collaboration, "Design, performance and future prospects of vertex detectors at the FCC-ee", *PoS* **ICHEP2024** (2025) 1062, `doi:10.22323/1.476.1062`, `arXiv:2506.02675`.

[48] G. Bencivenni et al., "The μ-RWELL layouts for high particle rate", *JINST* **14** (2019), no. 05, P05014, `doi:10.1088/1748-0221/14/05/P05014`, `arXiv:1903.11017`.

[49] N. Bacchetta et al., “CLD – A Detector Concept for the FCC-ee”, `arXiv:1911.12230`.

[50] G. Sadowski, J. Andrea, A. Besson, and Z. E. Bitar, “Tracking Performance Studies for Future Circular Collider (FCCee) with CLD Detector”, *EPJ Web Conf.* **315** (2024) 03003, `doi:10.1051/epjconf/202431503003`.

[51] J. Pekkanen, “ALLEGRO FCC-ee detector concept & Noble liquid calorimetry”, *Nucl. Instrum. Meth. A* **1069** (2024) 169921, `doi:10.1016/j.nima.2024.169921`.

[52] H. Abramowicz et al., “The International Linear Collider Technical Design Report - Volume 4: Detectors”, `arXiv:1306.6329`.

[53] H. Kolanoski and N. Wermes, “Particle Detectors: Fundamentals and Applications”. Oxford University Press, 2020. `doi:10.1093/oso/9780198858362.001.0001`, ISBN 9780198858362.

[54] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems”, *J. Fluids Eng.* **82** (1960), no. 1, 35–45, `doi:10.1115/1.3662552`.

[55] **CMS** Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018), no. 05, P05011, `doi:10.1088/1748-0221/13/05/P05011`, `arXiv:1712.07158`.

[56] T. Suehara and T. Tanabe, “LCFIPlus: A Framework for Jet Analysis in Linear Collider Studies”, *Nucl. Instrum. Meth. A* **808** (2016) 109–116, `doi:10.1016/j.nima.2015.11.054`, `arXiv:1506.08371`.

[57] **CMS** Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017), no. 10, P10003, `doi:10.1088/1748-0221/12/10/P10003`, `arXiv:1706.04965`.

[58] J. S. Marshall and M. A. Thomson, “Pandora Particle Flow Algorithm”, in *International Conference on Calorimetry for the High Energy Frontier*, pp. 305–315. 2013. `arXiv:1308.4537`.

[59] F. Bedeschi, L. Gouskos, and M. Selvaggi, “Jet flavour tagging for future colliders with fast simulation”, *Eur. Phys. J. C* **82** (2022), no. 7, 646, `doi:10.1140/epjc/s10052-022-10609-1`, `arXiv:2202.03285`.

[60] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, “Parton Fragmentation and String Dynamics”, *Phys. Rept.* **97** (1983) 31–145, `doi:10.1016/0370-1573(83)90080-7`.

[61] F. V. Tkachov, “Measuring multi - jet structure of hadronic energy flow or What is a jet?”, *Int. J. Mod. Phys. A* **12** (1997) 5411–5529, `doi:10.1142/S0217751X97002899`, `arXiv:hep-ph/9601308`.

[62] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-k_t jet clustering algorithm”, *JHEP* **04** (2008) 063, `doi:10.1088/1126-6708/2008/04/063`, `arXiv:0802.1189`.

[63] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J. C* **72** (2012) 1896, `doi:10.1140/epjc/s10052-012-1896-2`, `arXiv:1111.6097`.

[64] S. Catani et al., “New clustering algorithm for multi - jet cross-sections in e+ e- annihilation”, *Phys. Lett. B* **269** (1991) 432–438, `doi:10.1016/0370-2693(91)90196-W`.

[65] CMS Collaboration, "SWGuideBTagMCTools". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideBTagMCTools`, 2015. Accessed: 2025-07-25.

[66] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119–126, `doi:10.1016/j.physletb.2007.09.077`, `arXiv:0707.1378`.

[67] T. M. Mitchell, "Machine Learning". McGraw-Hill, 1997. ISBN 9780070428072.

[68] **LHCb** Collaboration, "Machine Learning based Global Particle Identification Algorithms at the LHCb Experiment", *EPJ Web Conf.* **214** (2019) 06011, `doi:10.1051/epjconf/201921406011`.

[69] S. Mondal and L. Mastrolorenzo, "Machine learning in high energy physics: a review of heavy-flavor jet tagging at the LHC", *Eur. Phys. J. ST* **233** (2024), no. 15-16, 2657–2686, `doi:10.1140/epjs/s11734-024-01234-y`, `arXiv:2404.01071`.

[70] P. Baldi, P. Sadowski, and D. Whiteson, "Searching for Exotic Particles in High-Energy Physics with Deep Learning", *Nature Commun.* **5** (2014) 4308, `doi:10.1038/ncomms5308`, `arXiv:1402.4735`.

[71] **ATLAS** Collaboration, "New techniques for jet calibration with the ATLAS detector", *Eur. Phys. J. C* **83** (2023), no. 8, 761, `doi:10.1140/epjc/s10052-023-11837-9`, `arXiv:2303.17312`.

[72] S. Cheong et al., "Parametrizing the Detector Response with Neural Networks", *JINST* **15** (2020), no. 01, P01030, `doi:10.1088/1748-0221/15/01/P01030`, `arXiv:1910.03773`.

[73] V. Belis, P. Odagiu, and T. K. Aarrestad, "Machine learning for anomaly detection in particle physics", *Rev. Phys.* **12** (2024) 100091, `doi:10.1016/j.revip.2024.100091`, `arXiv:2312.14190`.

[74] R. Kansal et al., "Evaluating generative models in high energy physics", *Phys. Rev. D* **107** (2023), no. 7, 076017, `doi:10.1103/PhysRevD.107.076017`, `arXiv:2211.10295`.

[75] X. Pang, S. Thulasidasan, and L. Rybarcyk, "Autonomous Control of a Particle Accelerator using Deep Reinforcement Learning", `arXiv:2010.08141`.

[76] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain", *Psychological Review* **65** (1958), no. 6, 386–408.

[77] A. Novikoff, "On Convergence Proofs on Perceptrons",.

[78] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Networks* **2** (1989), no. 5, 359–366, `doi:https://doi.org/10.1016/0893-6080(89)90020-8`.

[79] S. Hooker, "The Hardware Lottery", `https://arxiv.org/abs/2009.06489`, 2020.

[80] K. Chellapilla, S. Puri, and P. Simard, "High Performance Convolutional Neural Networks for Document Processing", in *Tenth International Workshop on Frontiers in Handwriting Recognition*, G. Lorette, ed., Université de Rennes 1. Suvisoft, La Baule (France), October, 2006. http://www.suvisoft.com.

[81] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, Big, Simple Neural Nets for Handwritten Digit Recognition", *Neural Computation* **22** (December,

2010) 3207–3220, `doi:10.1162/neco_a_00052`.

[82] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds., volume 25. Curran Associates, Inc., 2012.

[83] K. Rupp, “50 Years of Microprocessor Trend Data”. `https://github.com/karlrupp/microprocessor-trend-data`, 2022. Accessed: 2025-06-03.

[84] M. Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems”, `arXiv:1603.04467`.

[85] A. Paszke et al., “PyTorch: An Imperative Style, High-Performance Deep Learning Library”, `https://arxiv.org/abs/1912.01703`, 2019.

[86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors”, *Nature* **323** (1986) 533–536, `doi:10.1038/323533a0`.

[87] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, `https://arxiv.org/abs/1412.6980`, 2017.

[88] L. Liu et al., “On the Variance of the Adaptive Learning Rate and Beyond”, `https://arxiv.org/abs/1908.03265`, 2021.

[89] J. Gallicchio and M. D. Schwartz, “Quark and Gluon Jet Substructure”, *JHEP* **04** (2013) 090, `doi:10.1007/JHEP04(2013)090`, `arXiv:1211.7038`.

[90] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition”, *Proceedings of the IEEE* **86** (1998), no. 11, 2278–2324, `doi:10.1109/5.726791`.

[91] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, `https://arxiv.org/abs/1409.1556`, 2015.

[92] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning”, `https://arxiv.org/abs/1602.07261`, 2016.

[93] D. Guest, K. Cranmer, and D. Whiteson, “Deep Learning and its Application to LHC Physics”, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161–181, `doi:10.1146/annurev-nucl-101917-021019`, `arXiv:1806.11484`.

[94] H. Qu and L. Gouskos, “ParticleNet: Jet Tagging via Particle Clouds”, *Phys. Rev. D* **101** (2020), no. 5, 056019, `doi:10.1103/PhysRevD.101.056019`, `arXiv:1902.08570`.

[95] V. Mikuni and F. Canelli, “ABCNet: An attention-based method for particle tagging”, *Eur. Phys. J. Plus* **135** (2020), no. 6, 463, `doi:10.1140/epjp/s13360-020-00497-3`, `arXiv:2001.05311`.

[96] V. Mikuni and F. Canelli, “Point cloud transformers applied to collider physics”, *Mach. Learn. Sci. Tech.* **2** (2021), no. 3, 035027, `doi:10.1088/2632-2153/ac07f6`, `arXiv:2102.05073`.

[97] J. L. Elman, “Finding structure in time”, *Cognitive Science* **14** (1990), no. 2, 179–211,

`doi:10.1207/s15516709cog1402_1`.

[98] M. I. Jordan, “Serial order: A parallel distributed processing approach”, in *Neural Network Models of Cognition*, volume 121 of *Advances in Psychology*, pp. 471–495. Elsevier, 1997. `doi:10.1016/S0166-4115(97)80111-2`.

[99] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult”, *IEEE Transactions on Neural Networks* **5** (1994), no. 2, 157–166, `doi:10.1109/72.279181`.

[100] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, *Neural computation* **9** (1997), no. 8, 1735–1780.

[101] K. Cho et al., “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”, `https://arxiv.org/abs/1406.1078`, 2014.

[102] E. Bols et al., “Jet Flavour Classification Using DeepJet”, *JINST* **15** (2020), no. 12, P12012, `doi:10.1088/1748-0221/15/12/P12012`, `arXiv:2008.10519`.

[103] **ATLAS** Collaboration, “Identification of Jets Containing *b*-Hadrons with Recurrent Neural Networks at the ATLAS Experiment”,.

[104] **ATLAS** Collaboration, “Graph Neural Network Jet Flavour Tagging with the ATLAS Detector”,.

[105] H. Serviansky et al., “Set2Graph: Learning Graphs From Sets”, `arXiv:2002.08772`.

[106] J. Shlomi et al., “Secondary vertex finding in jets with neural networks”, *Eur. Phys. J. C* **81** (2021), no. 6, 540, `doi:10.1140/epjc/s10052-021-09342-y`, `arXiv:2008.02831`.

[107] J. Duarte and J.-R. Vlimant, “Graph Neural Networks for Particle Tracking and Reconstruction”, `doi:10.1142/9789811234033_0012`, `arXiv:2012.01249`.

[108] J. Pata et al., “MLPF: Efficient machine-learned particle-flow reconstruction using graph neural networks”, *Eur. Phys. J. C* **81** (2021), no. 5, 381, `doi:10.1140/epjc/s10052-021-09158-w`, `arXiv:2101.08578`.

[109] **Exa.TrkX** Collaboration, “Graph Neural Networks for Particle Reconstruction in High Energy Physics detectors”, in *33rd Annual Conference on Neural Information Processing Systems*. 3, 2020. `arXiv:2003.11603`.

[110] H. Qu, C. Li, and S. Qian, “Particle Transformer for Jet Tagging”, `arXiv:2202.03772`.

[111] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate”, `https://arxiv.org/abs/1409.0473`, 2016.

[112] K. Cho et al., “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation”, *arXiv preprint arXiv:1406.1078* (2014).

[113] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to Sequence Learning with Neural Networks”, in *Advances in Neural Information Processing Systems*, volume 27. 2014. `arXiv:1409.3215`.

[114] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate”, *arXiv preprint arXiv:1409.0473* (2014) `arXiv:1409.0473`.

[115] M.-T. Luong, H. Pham, and C. D. Manning, “Effective Approaches to Attention-based

Neural Machine Translation", *arXiv preprint arXiv:1508.04025* (2015) `arXiv:1508.04025`.

[116] A. Vaswani et al., "Attention Is All You Need", in *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. `arXiv:1706.03762`.

[117] **ATLAS** Collaboration, "Flavour tagging with graph neural networks with the ATLAS detector", in *30th International Workshop on Deep-Inelastic Scattering and Related Subjects*. 6, 2023. `arXiv:2306.04415`.

[118] L. Lonnblad, C. Peterson, and T. Rognvaldsson, "Using neural networks to identify jets", *Nucl. Phys. B* **349** (1991) 675–702, `doi:10.1016/0550-3213(91)90392-B`.

[119] T. Behnke and D. G. Charlton, "Electroweak measurements using heavy quarks at LEP", *Phys. Scripta* **52** (1995) 133–157, `doi:10.1088/0031-8949/52/2/002`.

[120] C. Bortoletto, A. De Angelis, and L. Lanceri, "Tagging the decays of the Z0 boson into b quark pairs with a neural network classifier", *Nucl. Instrum. Meth. A* **306** (1991) 459–466, `doi:10.1016/0168-9002(91)90039-S`.

[121] L. Bellantoni et al., "Using neural networks with jet shapes to identify b jets in e+ e- interactions", *Nucl. Instrum. Meth. A* **310** (1991) 618–622, `doi:10.1016/0168-9002(91)91108-8`.

[122] B. P. Roe et al., "Boosted decision trees, an alternative to artificial neural networks", *Nucl. Instrum. Meth. A* **543** (2005), no. 2-3, 577–584, `doi:10.1016/j.nima.2004.12.018`, `arXiv:physics/0408124`.

[123] H.-J. Yang, B. P. Roe, and J. Zhu, "Studies of boosted decision trees for MiniBooNE particle identification", *Nucl. Instrum. Meth. A* **555** (2005) 370–385, `doi:10.1016/j.nima.2005.09.022`, `arXiv:physics/0508045`.

[124] **D0** Collaboration, "Evidence for production of single top quarks and first direct measurement of |Vtb|", *Phys. Rev. Lett.* **98** (2007) 181802, `doi:10.1103/PhysRevLett.98.181802`, `arXiv:hep-ex/0612052`.

[125] **D0** Collaboration, "Evidence for production of single top quarks", *Phys. Rev. D* **78** (2008) 012005, `doi:10.1103/PhysRevD.78.012005`, `arXiv:0803.0739`.

[126] J. Bastos, "Tagging heavy flavours with boosted decision trees", `arXiv:physics/0702041`.

[127] **CMS** Collaboration, "Identification of b-Quark Jets with the CMS Experiment", *JINST* **8** (2013) P04013, `doi:10.1088/1748-0221/8/04/P04013`, `arXiv:1211.4462`.

[128] **ATLAS** Collaboration, "Performance of *b*-Jet Identification in the ATLAS Experiment", *JINST* **11** (2016), no. 04, P04008, `doi:10.1088/1748-0221/11/04/P04008`, `arXiv:1512.01094`.

[129] **ATLAS** Collaboration, "Secondary vertex finding for jet flavour identification with the ATLAS detector",.

[130] **CMS** Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2",.

[131] **ATLAS** Collaboration, "Calibration of the performance of *b*-tagging for *c* and

light-flavour jets in the 2012 ATLAS data",.

[132] **ATLAS** Collaboration, "Expected performance of the ATLAS b-tagging algorithms in Run-2",.

[133] **ATLAS** Collaboration, "Commissioning of the ATLAS b-tagging algorithms using $t\bar{t}$ events in early Run-2 data",.

[134] H. Luo et al., "Quark jet versus gluon jet: fully-connected neural networks with high-level features", *Sci. China Phys. Mech. Astron.* **62** (2019), no. 9, 991011, `doi:10.1007/s11433-019-9390-8`, `arXiv:1712.03634`.

[135] **CMS** Collaboration, "Heavy flavor identification at CMS with deep neural networks",.

[136] **ATLAS** Collaboration, "Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run",.

[137] **ATLAS** Collaboration, "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **79** (2019), no. 11, 970, `doi:10.1140/epjc/s10052-019-7450-8`, `arXiv:1907.05120`.

[138] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, "Deep learning in color: towards automated quark/gluon jet discrimination", *JHEP* **01** (2017) 110, `doi:10.1007/JHEP01(2017)110`, `arXiv:1612.01551`.

[139] J. S. H. Lee, I. Park, I. J. Watson, and S. Yang, "Quark-Gluon Jet Discrimination Using Convolutional Neural Networks", *J. Korean Phys. Soc.* **74** (2019), no. 3, 219–223, `doi:10.3938/jkps.74.219`, `arXiv:2012.02531`.

[140] **ATLAS** Collaboration, "Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector",.

[141] **CMS** Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", *JINST* **15** (2020), no. 06, P06005, `doi:10.1088/1748-0221/15/06/P06005`, `arXiv:2004.08262`.

[142] Y. Wang et al., "Dynamic Graph CNN for Learning on Point Clouds", *ACM Trans. Graph.* **38** (October, 2019) `doi:10.1145/3326362`.

[143] **CMS** Collaboration, "Adversarial training for b-tagging algorithms in CMS",.

[144] A. Stein et al., "Improving Robustness of Jet Tagging Algorithms with Adversarial Training", *Comput. Softw. Big Sci.* **6** (2022), no. 1, 15, `doi:10.1007/s41781-022-00087-1`, `arXiv:2203.13890`.

[145] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, "Weakly Supervised Classification in High Energy Physics", *JHEP* **05** (2017) 145, `doi:10.1007/JHEP05(2017)145`, `arXiv:1702.00414`.

[146] E. M. Metodiev, B. Nachman, and J. Thaler, "Classification without labels: Learning from mixed samples in high energy physics", *JHEP* **10** (2017) 174, `doi:10.1007/JHEP10(2017)174`, `arXiv:1708.02949`.

[147] A. Andreassen, I. Feige, C. Frye, and M. D. Schwartz, "JUNIPR: a Framework for Unsupervised Machine Learning in Particle Physics", *Eur. Phys. J. C* **79** (2019), no. 2, 102, `doi:10.1140/epjc/s10052-019-6607-9`, `arXiv:1804.09720`.

[148] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, “Learning to classify from impure samples with high-dimensional data”, *Phys. Rev. D* **98** (2018), no. 1, 011502, `doi:10.1103/PhysRevD.98.011502`, `arXiv:1801.10158`.

[149] E. Alvarez et al., “Exploring unsupervised top tagging using Bayesian inference”, *SciPost Phys. Core* **6** (2023) 046, `doi:10.21468/SciPostPhysCore.6.2.046`, `arXiv:2212.13583`.

[150] C. Bierlich et al., “A comprehensive guide to the physics and usage of PYTHIA 8.3”, *SciPost Phys. Codeb.* **2022** (2022) 8, `doi:10.21468/SciPostPhysCodeb.8`, `arXiv:2203.11601`.

[151] **DELPHES** Collaboration, “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`.

[152] F. Bedeschi, “A detector concept proposal for a circular e^+e^- collider”, *PoS* **ICHEP2020** (2021) 819, `doi:10.22323/1.390.0819`.

[153] M. Selvaggi, F. Bedeschi, and M. Ghilardi, “FCC-ee IDEA detector Delphes card”,.

[154] A. J. Larkoski, J. Thaler, and W. J. Waalewijn, “Gaining (Mutual) Information about Quark/Gluon Discrimination”, *JHEP* **11** (2014) 129, `doi:10.1007/JHEP11(2014)129`, `arXiv:1408.3122`.

[155] **CMS** Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”,.

[156] Y. Nakai, D. Shih, and S. Thomas, “Strange Jet Tagging”, `arXiv:2003.09517`.

[157] E. Nakano, “Belle PID”, *Nucl. Instrum. Meth. A* **494** (2002) 402–408, `doi:10.1016/S0168-9002(02)01510-3`.

[158] **Particle Data** Group, “Review of particle physics”, *Phys. Rev. D* **110** (2024), no. 3, 030001, `doi:10.1103/PhysRevD.110.030001`.

[159] K. Gautam, “Flavour Identification Techniques”, Master’s thesis, University of Copenhagen, 2020. Presented 18 Sep 2020.

[160] C. Helsens et al., “HEP-FCC/FCCAnalyses: v0.9.0”, `https://doi.org/10.5281/zenodo.10693709`, February, 2024. `doi:10.5281/zenodo.10693709`.

[161] F. Bedeschi, “A vertex fitting package”, `doi:10.17181/hvcpv-bk752`, `arXiv:2409.19326`.

[162] M. R. Zhang, J. Lucas, G. Hinton, and J. Ba, “Lookahead Optimizer: k steps forward, 1 step back”, `https://arxiv.org/abs/1907.08610`, 2019.

[163] **FCC** Collaboration, “Jet-Flavour Tagging at FCC-ee”, *PoS* **ICHEP2022** (2022) 1147, `doi:10.22323/1.414.1147`, `arXiv:2210.10322`.

[164] R. D. Field and R. P. Feynman, “A Parametrization of the Properties of Quark Jets”, *Nucl. Phys. B* **136** (1978) 1, `doi:10.1016/0550-3213(78)90015-9`.

[165] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, “Jet Charge at the LHC”, *Phys. Rev. Lett.* **110** (2013), no. 21, 212001, `doi:10.1103/PhysRevLett.110.212001`,

`arXiv:1209.2421`.

[166] P. Azzi, L. Gouskos, M. Selvaggi, and F. Simon, "Higgs and top physics reconstruction challenges and opportunities at FCC-ee", *Eur. Phys. J. Plus* **137** (2022), no. 1, 39, `doi:10.1140/epjp/s13360-021-02223-z`, `arXiv:2107.05003`.

[167] A. Albert et al., "Strange quark as a probe for new physics in the Higgs sector", in *Snowmass 2021.* 3, 2022. `arXiv:2203.07535`.

[168] L. Breiman, "Random Forests", *Mach. Learn.* **45** (2001), no. 1, 5–32, `doi:10.1023/A:1010933404324`.

[169] A. Fisher, C. Rudin, and F. Dominici, "All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously", `https://arxiv.org/abs/1801.01489`, 2019.

[170] Glen Cowan, Eilam Gross, "Discovery significance with statistical uncertainty in the background estimate", ATLAS Statistics Forum, 2008.

[171] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, `doi:10.1140/epjc/s10052-011-1554-0`, `arXiv:1007.1727`.

[172] H. Spieler, "Semiconductor Detector Systems", volume v.12 of *Semiconductor Science and Technology.* Oxford University Press, Oxford, 2005. ISBN 978-0-19-852784-8.

[173] G. Knoll, "Radiation Detection and Measurement (4th ed.)". John Wiley, Hoboken, NJ, 2010. ISBN 978-0-470-13148-0.

[174] G. Lutz, "Semiconductor Radiation Detectors: Device Physics". Springer, New York, 1999. ISBN 978-3-540-64859-8.

[175] H. Bethe, "Theory of the Passage of Fast Corpuscular Rays Through Matter", *Annalen Phys.* **5** (1930) 325–400, `doi:10.1002/andp.19303970303`.

[176] M. Friedl, "The CMS silicon strip tracker and its electronic readout". PhD thesis, Technische U. Wien, 2001.

[177] L. D. Landau, "On the Energy Loss of Fast Particles by Ionisation", *J. Phys. (USSR)* **8** (1944) 201–205, `doi:10.1016/B978-0-08-010586-4.50061-4`.

[178] P. V. Vavilov, "Ionization Losses of High-Energy Heavy Particles", *Sov. Phys. JETP* **5** (1957) 749–751.

[179] H. Bichsel, "Straggling of Heavy Charged Particles: Comparison of Born Hydrogenic-Wave-Function Approximation with Free-Electron Approximation", *Phys. Rev. B* **1** (April, 1970) 2854–2862, `doi:10.1103/PhysRevB.1.2854`.

[180] S. Meroli, D. Passeri, and L. Servoli, "Energy loss measurement for charged particles in very thin silicon layers", *JINST* **6** (2011) P06013, `doi:10.1088/1748-0221/6/06/P06013`.

[181] B. Rossi and K. Greisen, "Cosmic-ray theory", *Rev. Mod. Phys.* **13** (1941) 240–309, `doi:10.1103/RevModPhys.13.240`.

[182] J. Cabello and K. Wells, "The spatial resolution of silicon-based electron detectors in beta-autoradiography", *Phys. Med. Biol.* **55** (2010), no. 6, 1677–1699,

`doi:10.1088/0031-9155/55/6/010`.

[183] W. Shockley, “Currents to conductors induced by a moving point charge”, *J. Appl. Phys.* **9** (1938), no. 10, 635–636, `doi:10.1063/1.1710367`.

[184] S. Ramo, “Currents induced by electron motion”, *Proc. Ire.* **27** (1939) 584–585, `doi:10.1109/JRPROC.1939.228757`.

[185] **ALICE** Collaboration, “ALPIDE, the Monolithic Active Pixel Sensor for the ALICE ITS upgrade”, *Nucl. Instrum. Meth. A* **824** (2016) 434–438, `doi:10.1016/j.nima.2015.09.057`.

[186] T.-H. Lee et al., “Analysis of 1/f noise in CMOS preamplifier with CDS circuit”, *IEEE Trans. Nucl. Sci.* **49** (2002), no. 4, 1819–1823, `doi:10.1109/TNS.2002.801514`.

[187] I. Peric, “A novel monolithic pixelated particle detector implemented in high-voltage CMOS technology”, *Nucl. Instrum. Meth. A* **582** (2007) 876–885, `doi:10.1016/j.nima.2007.07.115`.

[188] N. Barchetta, P. Collins, and P. Riedler, “Tracking and vertex detectors at FCC-ee”, *Eur. Phys. J. Plus* **137** (2022), no. 2, 231, `doi:10.1140/epjp/s13360-021-02323-w`, `arXiv:2112.13019`.

[189] **CLICdp** Collaboration, “Design of a monolithic HR-CMOS sensor chip for the CLIC silicon tracker”, *PoS* **TWEPP2018** (2019) 072, `doi:10.22323/1.343.0072`.

[190] L. Zhang, “Development of a CMOS pixel sensor for the outer layers of the ILC vertex detector”. Phd thesis, Université de Strasbourg, 2013. NNT : 2013STRAE036.

[191] Y. Zhang et al., “Development of a CMOS pixel sensor prototype for the high hit rate CEPC vertex detector”, *Nucl. Instrum. Meth. A* **1042** (2022) 167442, `doi:10.1016/j.nima.2022.167442`.

[192] **ALICE** Collaboration, “Technical Design Report for the Upgrade of the ALICE Inner Tracking System”, *J. Phys. G* **41** (2014) 087002, `doi:10.1088/0954-3899/41/8/087002`.

[193] **ALICE** Collaboration, “Upgrade of the ALICE ITS detector”, *Nucl. Instrum. Meth. A* **1032** (2022) 166632, `doi:10.1016/j.nima.2022.166632`, `arXiv:2111.08301`.

[194] G. Contin et al., “The STAR MAPS-based PiXeL detector”, *Nucl. Instrum. Meth. A* **907** (2018) 60–80, `doi:10.1016/j.nima.2018.03.003`, `arXiv:1710.02176`.

[195] E. D. R. R. P. Group, “The 2021 ECFA detector research and development roadmap”. CERN, Geneva, 10, 2021. `doi:10.17181/CERN.XDPL.W2EX`.

[196] **ALICE** Collaboration, “Performance studies of the CE-65v2 MAPS prototype structure”, *JINST* **20** (2025), no. 03, C03033, `doi:10.1088/1748-0221/20/03/C03033`, `arXiv:2502.04070`.

[197] G. A. Rinella et al., “Characterization of analogue Monolithic Active Pixel Sensor test structures implemented in a 65 nm CMOS imaging process”, *Nucl. Instrum. Meth. A* **1069** (2024) 169896, `doi:10.1016/j.nima.2024.169896`, `arXiv:2403.08952`.

[198] G. A. Rinella et al., “Digital pixel test structures implemented in a 65 nm CMOS process”, *Nucl. Instrum. Meth. A* **1056** (2023) 168589,

`doi:10.1016/j.nima.2023.168589`, `arXiv:2212.08621`.

[199] S. Bugiel et al., "Charge sensing properties of monolithic CMOS pixel sensors fabricated in a 65 nm technology", *Nucl. Instrum. Meth. A* **1040** (2022) 167213, `doi:10.1016/j.nima.2022.167213`.

[200] **ALICE** Collaboration, "First measurements with monolithic active pixel test structures produced in a 65 nm CMOS process", *JINST* **19** (2024), no. 02, C02017, `doi:10.1088/1748-0221/19/02/C02017`, `arXiv:2309.14814`.

[201] **ALICE** Collaboration, "Characterisation of analogue MAPS fabricated in 65 nm technology for the ALICE ITS3", *Nucl. Instrum. Meth. A* **1068** (2024) 169787, `doi:10.1016/j.nima.2024.169787`, `arXiv:2409.07543`.

[202] G. A. Rinella et al., "Time performance of Analog Pixel Test Structures with in-chip operational amplifier implemented in 65 nm CMOS imaging process", *Nucl. Instrum. Meth. A* **1070** (2025) 170034, `doi:10.1016/j.nima.2024.170034`, `arXiv:2407.18528`.

[203] A. Simancas et al., "Simulations and performance studies of a MAPS in 65 nm CMOS imaging technology", *Nucl. Instrum. Meth. A* **1064** (2024) 169414, `doi:10.1016/j.nima.2024.169414`, `arXiv:2402.14524`.

[204] G. A. Rinella et al., "Further Characterisation of Digital Pixel Test Structures Implemented in a 65 nm CMOS Process", `arXiv:2505.05867`.

[205] W. Snoeys et al., "A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance", *Nucl. Instrum. Meth. A* **871** (2017) 90–96, `doi:10.1016/j.nima.2017.07.046`.

[206] M. Munker et al., "Simulations of CMOS pixel sensors with a small collection electrode, improved for a faster charge collection and increased radiation tolerance", *JINST* **14** (2019), no. 05, C05013, `doi:10.1088/1748-0221/14/05/C05013`, `arXiv:1903.10190`.

[207] A. Dorokhov, "Optimization of amplifiers for monolithic active pixel sensors", `doi:10.5170/CERN-2007-007.423`.

[208] S. Bugiel, A. Dorokhov, H. Pham, and I. Valin, "CE65 A/B/C/D Chip Documentation", Technical Report Version 1.1, Institut Pluridisciplinaire Hubert Curien (IPHC), 2021. Submitted November 2020, last updated March 2, 2021. Publicly redistributable.

[209] T. Kugathasan et al., "Monolithic CMOS sensors for sub-nanosecond timing", *Nucl. Instrum. Meth. A* **979** (2020) 164461, `doi:10.1016/j.nima.2020.164461`.

[210] **ALICE** Collaboration, "A readout system for monolithic pixel sensor prototypes towards the upgrade of the ALICE Inner Tracking System", *JINST* **18** (2023), no. 01, C01047, `doi:10.1088/1748-0221/18/01/C01047`.

[211] X. Mougeot et al., "Evaluations of the decay data of 55Fe, 55Co, 103mRh, 103Pd, 129mSn and 166Ho from the Decay Data Evaluation Project (DDEP)—2024", *Metrologia* **62** (feb, 2025) 029001, `doi:10.1088/1681-7575/adb275`.

[212] M. Suljic, "Study of Monolithic Active Pixel Sensors for the Upgrade of the ALICE Inner Tracking System". PhD thesis, University of Trieste, 2018. CERN-THESIS-2017-304.

[213] J. D. Berst et al., "Monolithic active pixel sensors for high resolution vertex detectors", in

6th Workshop on Electronics for LHC Experiments, pp. 535–539. 9, 2000.

[214] B. McCleary, “Digital imaging system testing and design using physical sensor characteristics”. PhD thesis, Univ. Southern California, 2009.

[215] D. Banerjee et al., “The North Experimental Area at the Cern Super Proton Synchrotron”, `doi:10.17181/CERN.GP3K.0S1Y`. Dedicated to Giorgio Brianti on the 50th anniversary of his founding the SPS Experimental Areas Group of CERN-Lab II and hence initiating the present Enterprise.

[216] R. Diener et al., “The DESY II Test Beam Facility”, *Nucl. Instrum. Meth. A* **922** (2019) 265–286, `doi:10.1016/j.nima.2018.11.133`, `arXiv:1807.09328`.

[217] M. Williams et al., “Corryvreckan - A Modular 4D Track Reconstruction and Analysis Software for Test Beam Data”, `https://doi.org/10.5281/zenodo.4384186`, December, 2020. `doi:10.5281/zenodo.4384186`.

[218] V. Blobel, “A new fast track-fit algorithm based on broken lines”, *Nucl. Instrum. Meth. A* **566** (2006) 14–17, `doi:10.1016/j.nima.2006.05.156`.

[219] C. Kleinwort, “General Broken Lines as advanced track fitting method”, *Nucl. Instrum. Meth. A* **673** (2012) 107–110, `doi:10.1016/j.nima.2012.01.024`, `arXiv:1201.4320`.

[220] C. J. Clopper and E. S. Pearson, “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial”, *Biometrika* **26** (1934), no. 4, 404–413, `doi:10.1093/biomet/26.4.404`.

[221] M. Mager, “Telescope Optimizer”. `https://mmager.web.cern.ch/telescope/tracking.html`, 2016. Accessed: September 16, 2024.

[222] A. Papoulis and S. U. Pillai, “Probability, Random Variables, and Stochastic Processes”. McGraw-Hill, Boston, 4th edition, 2002. ISBN 9780073660110.

[223] **ALICE** Collaboration, “Characterization of ALPIDE silicon sensors with inclined tracks”, *PoS* **ICHEP2020** (2021) 805, `doi:10.22323/1.390.0805`.

[224] “PETRA III: A low emittance synchrotron radiation source. Technical design report”,.

[225] S. Spannagel, “Test Beam Measurements for the Upgrade of the CMS Pixel Detector and Measurement of the Top Quark Mass from Differential Cross Sections”. PhD thesis, Hamburg U., 2016. DESY-THESIS-2016-010, CMS-TS-2016-010, CERN-THESIS-2016-059, `doi:10.3204/DESY-THESIS-2016-010`.

[226] A. Ilg, A. Macchiolo, and F. Palla, “The vertexing challenge at FCC-ee”, in *11th International Workshop on Semiconductor Pixel Detectors for Particles and Imaging*. 2, 2025. `arXiv:2502.04071`.

[227] “OCTOPUS Project – Optimized CMOS Technology fOr Precision in Ultra-thin Silicon”. `https://octopus.web.cern.ch/`.

[228] J. Soudier et al., “A versatile and fast pixel matrix read-out architecture for MAPS”, *Nucl. Instrum. Meth. A* **1067** (2024) 169663, `doi:10.1016/j.nima.2024.169663`.